# SIGNAL PROCESSING, HUMAN FACTORS, AND MODELLING TO SUPPORT BEDSIDE CARE IN THE INTENSIVE CARE UNIT

by

Matthias Görges

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Bioengineering

The University of Utah

August 2010

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

# STATEMENT OF DISSERTATION APPROVAL

The dissertation of Matthias Görges
has been approved by the following supervisory committee members:

| | | |
|---|---|---|
| Dwayne R. Westenskow | , Chair | 4/26/2010 |
| | | Date Approved |
| Douglas A. Christensen | , Member | 4/28/2010 |
| | | Date Approved |
| Robert S. MacLeod | , Member | 4/26/2010 |
| | | Date Approved |
| Boaz A. Markewitz | , Member | 4/26/2010 |
| | | Date Approved |
| Joseph A. Orr | , Member | 4/26/2010 |
| | | Date Approved |

and by Richard D. Rabbitt, Chair of
the Department of Bioengineering

and by Charles A. Wight, Dean of The Graduate School.

# ABSTRACT

Medical error causes preventable death in nearly 100,000 patients per year in the US alone. Common sources for error include medication related problems, technical equipment failure, interruptions, complicated and error-prone devices, information overload (providing too much patient data for one person to process effectively), and environmental problems like inadequate lighting or distracting ambient noise.

Intensive care units are one of the riskiest locations in a hospital, with up to 9 reported events per 100 patient days. This risk is in large contrast to anesthesia in the operating rooms. Here much advancement in the area of patient safety has been made in the past, dropping the average risk for anesthesia related death to less than 1 in 200,000 anesthetics—an improvement by a factor of 20 in the past 30 years. Improvements in technology and other innovations contributing to this success now need to be adapted for and implemented in the intensive care unit setting.

Nurses are increasingly regarded as key decision makers within the healthcare team, as they outnumber physicians 4:1. Reducing nurses' workload and improving medical decision making by providing decision support tools can have a significant impact in reducing the chances of medical errors.

This dissertation consists of four manuscripts: 1) a review of previous medical display evaluations, providing insight into solutions that have worked in the past; 2) a study on reducing false alarms and increasing the usefulness of the remaining alarms by introducing alarm delays and detecting alarm context, such as suctioning automatically silencing ventilator alarms; 3) a study of simplifying the frequent but complicated task of titrating vasoactive medications by providing a titration support tool that predicts blood pressure changes 5 minutes into the future; and 4) a study on supporting the triage of unfamiliar patients by introducing a far-view display that incorporates information from previously disparate devices and presents trend and alarm information at one easy to scan and interpret location.

My parents.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

I would first like to thank Dwayne Westenskow, Director of the Anesthesiology Bioengineering division, for inviting me to join his research group and providing me the opportunity to perform the research described in this dissertation. His expert advice, encouraging comments, integrity and dedication to research were invaluable. Thanks to Joseph Orr, Co-Director of the Anesthesiology Bioengineering division, for his guidance and support of my research activities in the lab, as well as providing me with the opportunity to be involved in a second display design and evaluation project during my time in Utah.

I appreciate the time and guidance provided by the members of my PhD supervisory committee: Dwayne R. Westenskow, Douglas A. Christensen, Robert S. MacLeod, Boaz A. Markewitz, and Joseph A. Orr. Their commitment to this project helped focus the project, and create a good balance between science, engineering and medicine.

Thanks to Kai Kück for having the foresight to start exploring context awareness and its applications in medicine. His support and encouraging feedback during our bimonthly conference calls helped keep the project on track and focused on the importance of applying engineering to medicine in order to improve patient care and safety. Without his involvement I would never have pursued this research.

I am also indebted to Mark Ansermino at the University of British Columbia, who reviewed this dissertation and provided many useful comments. I am looking forward to joining his research group as a postdoctoral fellow this fall.

I would like to express many thanks to other collaborators in this project: Jim Agutter for instructions in good design and superb suggestions and feedback during the creation of the far-view display. Nancy Staggers for providing me with the opportunity of expanding a project I started during her Human Factors class, which culminated in a review of existing monitoring evaluation literature. Sven Koch, not

only for working on the close-view nursing display, which is intended to accompany the far-view display developed, but also for frequent feedback in the far-view display design and evaluation process.

I am most thankful to Cris LaPierre and Lara Brewer for sharing the entire PhD experience with me - while becoming close friends - and frequently providing tips, encouragements and support for my research. I am thankful for the support of the Anesthesia department staff and my fellow labmates Tammy Anderson, Sören Höhne, Cameron Jacobson, and Carl Tams. Special thanks to David Liu for statistics and study design advice and Bryce Hill for providing exciting community projects totally unrelated to my research.

Special gratitude goes to my parents, whom this dissertation is dedicated to, my siblings and my friends for their support and encouragement during my long time away from home.

Permission to reprint the paper published in Anesthesia & Analgesia was granted by the International Anesthesia Research Society and the papers published in the Journal of Clinical Monitoring and Computing were granted by Springer Science + Business Media. Finally, I would like to thank Drägerwerk AG, Lübeck, Germany for their interest and support of this research project.

# CHAPTER 1

# INTRODUCTION

This dissertation is the compilation of my work at the University of Utah focusing on reducing nurses' workload and improving medical decision making, thereby reducing the chances of medical errors. It consists of four manuscripts. The first is a review of previous medical display evaluations. The remaining three are studies suggesting the following improvements to nurses' work: a) reduction of false alarms and increased usefulness of remaining alarms; b) simplification of a common but complicated task, titration of vasoactive medications; and c) support for triaging unfamiliar patients using a far-view display.

## 1.1 Background

### 1.1.1 Medical Error

In 1997 an Institute of Medicine report estimated the number of preventable deaths caused by medical error to be between 44,000 and 98,000.[1] This report started the modern patient-safety movement. Preventable medication errors have been found to occur in up to 1.5% of all hospital admissions.[2] Medical errors are common in intensive care units (ICUs), with 36-89 reported events per 1,000 ICU patient days.[3, 4] Causes of errors include complicated and error-prone devices, information overload (providing too much patient data for one person to process effectively), and environmental problems like inadequate lighting or distracting ambient noise.[5] The most common medical errors in the ICU are medication errors, problems with intravenous infusions, and technical equipment failure.[6] Problems in patient identification,[7] wrong patients or wrong location in operations,[8] interruptions,[9] and team communication in the operating room[10] are only some of the areas where improvements are needed and have been proposed. Computerized physician-order-entry or decision-support systems

can reduce certain types of medication error but have the drawbacks of slowing clinical workflow and introducing new errors if not performed carefully.[11]

### 1.1.2 Medical Decision Making

Evidence-based medicine[12, 13] aims to address the problem of clinical-practice variation by replacing personal clinical experience as the primary resource for medical decision-making with practice recommendations and guidelines based on systematic studies of populations.[14] Sources of medical decision-making support[15] include artificial neural networks,[16, 17] statistical methods such as Bayesian interference[18] or fuzzy logic,[19] case-based reasoning,[20] and expert systems.[21, 22] Data integration, using clinical dashboards[23] or single indicators combining multiple variables,[24] has shown promise for improving patient care.

Nurses are increasingly regarded as key decision makers within the healthcare team[25] and outnumber physicians 4:1. Nurses prefer humans as information sources as these deliver context specific information when needed. Additionally, literature use almost never occurs at the point of decision-making but rather after the fact.[25] Research information needs to be presented in formats maximized for limited consumption opportunities, as nurses have limited time to explore literature.[26] Finally, the follow-up report to "to err is human"[1] specifically asked for decision support tools, such as reminders and alerts.[27]

### 1.1.3 Human Factors

Human factors, the science of applying understanding of human capabilities and limitations to the design, development, and deployment of systems and services, has led to major safety improvements in aviation[28] and nuclear engineering.[29] More recently it has been applied to medicine, starting as early as the 1980s in the field of anesthesiology.[1] In this field, collections of preventable incidents[30] or closed insurance claims[31] led to recommendations for preventing and detecting such incidents.

Lack of situational awareness or inadequate situational awareness has been identified as one of the primary factors in accidents attributed to human error.[32] There are three levels of situational awareness: 1) perception, which includes detection of elements or identification of values; 2) comprehension, which includes the synthesis

of multiple elements towards understanding the current situation; and 3) projection, which extrapolates trends forward in time (e.g., for therapy planning). All three levels of situational awareness must be fulfilled to prevent errors.[33–35]

In the ICU, human factors techniques such as qualitative observations have been used to identify problems in commonly occurring tasks; for example, interruptions to a nurse's attention during medication preparation and tasks being forgotten because of large cognitive workload of nurses.[36] Safety problems caused by shortcomings in nontechnical skills such as task management, teamwork, situation awareness, and decision making can be analyzed using root-cause analysis or observational studies.[37] Clinical technologies such as graphical displays, medical-design interfaces and clinical-application designs have been analyzed for their usability and improvements have been reported, but they still need to focus more on nurses as their users.[38]

## 1.2 Goals and Contributions to the Literature

### 1.2.1 Motivation for Focusing on the Intensive Care Unit

Anesthesiology has been at the forefront of technology and patient safety, as practitioners of anesthesiology are enthusiastic about technological innovation.[39] Examples of innovation in this field include the introduction of cardiac monitoring, pulse oximetry, and capnography and have led to anesthesiology being acknowledged as a model for patient safety in medicine.[40] These technological improvements and other innovations now need to be adapted for and implemented in the ICU. The following four chapters contain manuscripts focusing on reducing nurses' workload and improving medical decision making, thereby reducing the chances of medical errors.

### 1.2.2 Review of Physiologic Monitoring Display Evaluations

The purpose of this evaluation, which forms Chapter 2 of this dissertation, was to present the findings of past physiologic monitoring display evaluations that demonstrate reductions in medical errors and provider workload (both physical and mental) and improvements in medical decision making. It provides an opportunity to examine past work across studies and learn which ideas worked well and which did not, and

it sets the stage for the design and conduct of future evaluations in two subsequent studies performed in this dissertation. Participants were faster detecting an adverse event or making a diagnosis or decision in 57% of the evaluations. They showed an improved accuracy in a clinical decision or diagnosis 67% of the studies measuring this and a perceived workload decrease in 43% of the studies accessing this variable. The majority of the evaluations (61%) used anesthesiologists, practitioners in a field from which many medical innovations originate, and only 16% used nurses. This highlights the need for future clinical studies to focus on participants besides anesthesiologists.

### 1.2.3 Alarm Reductions Using Delays and Clinical Context

The purpose of this study, which comprises Chapter 3 of this dissertation, was to identify methods for reducing the number of false alarms by using time delays and the correlations between alarms and clinical context. This information was obtained by observing health care providers caring for patients in the MICU. The study proposed a 19 sec alarm delay, which would have reduced 67% of the ignored and ineffective alarms, thereby reducing the noise level in the unit and potentially reducing nurses' workload. It identified nurses as the main monitoring users, making 66% of all visits to a patient's room, which should lead future research to design displays supporting nurses specifically. It also observed that nurses used equipment functions in a way not intended by the manufacturers (e.g., intentionally entering a smaller infusate volume than was available, so that the infusion pump alarm reminded them when the pump was nearly empty). These behaviors lead to unnecessary alarms. Additionally, nurses had to integrate information from many disparate sources, with only information from the cardiac monitor being available outside the patient's room. Finally, we observed that the titration of vasoactive medications was a challenging task, requiring significant nursing resources (in terms of staff availability as well as mental workload for the nurse performing this task). Future work should allow for combining clinical context, such as provider presence, performed tasks (suctioning causing alarm silencing, or titrating medications with predictions of vitals sign changes), and the patient's state in the physiological monitor.

### 1.2.4 Titration Advisory System with Patient Specific Sensitivity Identification

Chapter 4 of this dissertation is the first example of supporting nurses in their clinical practice, by reducing their workload and improving their decision making. The purpose of this study was to use simulation to test the feasibility of using small-step changes in infusion rates to automatically identify a patient's sensitivity to sodium-nitroprusside (SNP), dobutamine, or dopamine as the drug is being infused and to evaluate whether an advisory system that predicts blood pressure values 5 min in the future enhances a clinician's ability to manage SNP infusion. Findings indicate a 52-82% improvement in the accuracy of the mean arterial blood pressure (MAP) prediction when using the identification system for the three investigated medications (SNP, dopamine and dobutamine); a median time reduction of 6.1 min to reach the desired MAP; and a significant reduction of mental workload and effort. Finally, the sensitivity identification led to a proposed extension of existing therapy support indicators, such as the inspired oxygen fraction and ventilator provided minute volume supporting blood oxygen saturation, to vasoactive drugs altering heart rate, blood pressure or cardiac output.

### 1.2.5 Intensive Care Unit Far-View Display Supporting Triaging Tasks

Chapter 5 of this dissertation is the second example of supporting nurses in their clinical practice, by supporting them in triaging unfamiliar patients. The goal of the study was to test two hypotheses: a) the information provided by a far-view display allows a clinician to faster identify which patients need the most immediate attention, and b) the far-view display will reduce the clinicians' mental workload and improve situational awareness. The novel display was designed specifically for nurses as its main users (proposed in the Chapter 2) and includes infusion pumps indicating the time until they are empty (proposed in Chapter 3), as well as therapy support indicators (proposed in Chapter 4). It might find a future application not only in making triage decisions of unfamiliar patients but also in communicating patients' vital signs in change-of-shift reports. A nurse-specific close-view display, integrating multiple devices, such as cardiac patient monitors, infusion pumps, ventilators and

the electronic medical record, into a single easy to use device for nurses was designed and evaluated as a separate project performed by Sven Koch.[41]

## 1.3 References


1. Committee on Quality of Health Care in America. Errors in Health Care: A Leading Cause of Death and Injury. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err Is Human: Building a Safer Health System. 1st ed. National Academies Press; 2000. p. 26–48.

2. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA. 1997 Jan;277(4):307–311.

3. Osmon S, Harris CB, Dunagan WC, Prentice D, Fraser VJ, Kollef MH. Reporting of medical errors: an intensive care unit experience. Crit Care Med. 2004 Mar;32(3):727–733.

4. Rothschild JM, Landrigan CP, Cronin JW, Kaushal R, Lockley SW, Burdick E, et al. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. Crit Care Med. 2005 Aug;33(8):1694–1700.

5. Donchin Y, Seagull FJ. The hostile environment of the intensive care unit. Curr Opin Crit Care. 2002 Aug;8(4):316–320.

6. Flaatten H, Hevroy O. Errors in the intensive care unit (ICU). Experiences with an anonymous registration. Acta Anaesthesiol Scand. 1999 Jul;43(6):614–617.

7. Murphy MF, Kay JDS. Patient identification: problems and potential solutions. Vox Sang. 2004 Jul;87 Suppl 2:197–202.

8. Sandberg WS, Häkkinen M, Egan M, Curran PK, Fairbrother P, Choquette K, et al. Automatic detection and notification of 'wrong patient-wrong location' errors in the operating room. Surg Innov. 2005 Sep;12(3):253–260.

9. Liu D, Grundgeiger T, Sanderson PM, Jenkins SA, Leane TA. Interruptions and blood transfusion checks: lessons from the simulated operating room. Anesth Analg. 2009 Jan;108(1):219–222.

10. Davies JM. Team communication in the operating room. Acta Anaesthesiol Scand. 2005 Aug;49(7):898–901.

11. Handler JA, Feied CF, Coonan K, Vozenilek J, Gillam M, Peacock PRJ, et al. Computerized physician order entry and online decision support. Acad Emerg Med. 2004 Nov;11(11):1135–1141.

12. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. Health Aff (Millwood). 2005 Jan;24(1):18–28.

13. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 Jan;312(7023):71–72.

14. Eddy DM. Clinical Decision Making: From Theory to Practice: A Collection of Essays From the Journal of the American Medical Association. 1st ed. Sudbury, MA: Jones and Bartlett Publishers; 1996.

15. Berner ES, editor. Clinical Decision Support Systems: Theory and Practice. 2nd ed. New York, NY: Springer-Verlag; 2006.

16. Kattan MW, Beck JR. Artificial neural networks for medical classification decisions. Arch Pathol Lab Med. 1995 Aug;119(8):672–677.

17. Sharpe PK, Caleb P. Artificial neural networks within medical decision support systems. Scand J Clin Lab Invest Suppl. 1994;219:3–11.

18. Ashby D. Bayesian statistics in medicine: a 25 year review. Stat Med. 2006 Nov;25(21):3589–3631.

19. Bates JHT, Young MP. Applying fuzzy logic to medical decision making in the intensive care unit. Am J Respir Crit Care Med. 2003 Apr;167(7):948–952.

20. Dussart C, Pommier P, Siranyan V, Grelaud G, Dussart S. Optimizing clinical practice with case-based reasoning approach. J Eval Clin Pract. 2008 Oct;14(5):718–720.

21. Brokel JM, Shaw MG, Nicholson C. Expert clinical rules automate steps in delivering evidence-based care in the electronic health record. Comput Inform Nurs. 2006 Jul;24(4):196–205.

22. Heldt T, Long B, Verghese GC, Szolovits P, Mark RG. Integrating data, models, and reasoning in critical care. Conf Proc IEEE Eng Med Biol Soc. 2006;1:350–353.

23. Egan M. Clinical dashboards: impact on workflow, care quality, and patient safety. Crit Care Nurs Q. 2006 Oct;29(4):354–361.

24. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth. 2006 Jul;97(1):64–68.

25. Thompson C, Cullum N, McCaughan D, Sheldon T, Raynor P. Nurses, information use, and clinical decision making–the real world potential for evidence-based decisions in nursing. Evid Based Nurs. 2004 Jul;7(3):68–72.

26. Thompson C, McCaughan D, Cullum N, Sheldon T, Raynor P. Barriers to evidence-based practice in primary care nursing–why viewing decision-making as context is helpful. J Adv Nurs. 2005 Nov;52(4):432–444.

27. Committee on Data Standards for Patient Safety. Patient Safety: Achieving a New Standard for Care. Aspden P, Corrigan JM, Wolcott J, Erickson S, editors. Washington, DC: National Academies Press; 2004.

28. Stokes A, Wickens C. Aviation Displays. In: Wiener E, Nagel D, editors. Human Factors in Aviation. San Diego, CA: Academic Press; 1988. p. 387–431.

29. Stanton NA, editor. Human Factors in Nuclear Safety. Bristol, PA: Taylor & Francis; 1996.

30. Cooper JB, Newbower RS, Kitz RJ. An analysis of major errors and equipment failures in anesthesia management: considerations for prevention and detection. Anesthesiology. 1984 Jan;60(1):34–42.

31. Cheney FW. The American Society of Anesthesiologists Closed Claims Project: what have we learned, how has it affected practice, and how will it affect practice in the future? Anesthesiology. 1999 Aug;91(2):552–556.

32. Nullmeyer RT, Stella D, Montijo GA, Harden SW. Human Factors in Air Force Flight Mishaps: Implications for Change. In: The Interservice/Industry Training, Simulation & Education Conference. 2260. Arlington, VA; 2005. p. 1–11.

33. Drews F, Westenskow D. Human computer interaction in health care. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Lawrence Erlbaum Associates; 2006. p. 423–38.

34. Goodstein L. Discriminative display support for process operators. In: Rasmussen J, Rouse W, editors. Human Detection and Diagnosis of System Failures. Springer; 1981. p. 433–449.

35. Pew RW. The State of Situation Awareness Measurement: Heading Toward the Next Century, in Situation Awareness Analysis and Measurement. In: Endsley MR, Garland DJ, editors. Situation Awareness Analysis and Measurement. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. p. 33–47.

36. Potter P, Wolf L, Boxerman S, Grayson D, Sledge J, Dunagan C, et al. Understanding the cognitive work of nursing in the acute care environment. J Nurs Adm. 2005 Jul;35(7-8):327–335.

37. Reader T, Flin R, Lauche K, Cuthbertson BH. Non-technical skills in the intensive care unit. Br J Anaesth. 2006 May;96(5):551–559.

38. Alexander G, Staggers N. A systematic review of the designs of clinical technology: findings and recommendations for future research. ANS Adv Nurs Sci. 2009 Jul;32(3):252–279.

39. Melo MFV, Leone BJ. Introduction of new monitors into clinical anesthesia. Anesth Analg. 2008 Sep;107(3):749–750.

40. Gaba DM. Anaesthesiology as a model for patient safety in health care. BMJ. 2000 Mar;320(7237):785–788.

41. Koch SH, Staggers N, Weir CR, Agutter J, Liu D, Westenskow DR. Integrated Information Displays for ICU Nurses: Field Observations, Display Design, and Display Evaluation. In: Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society. 429. San Francisco, CA; 2010. .

# CHAPTER 2

# EVALUATIONS OF PHYSIOLOGIC MONITORING DISPLAYS: A SYSTEMATIC REVIEW*

## 2.1 Abstract

The purpose of this paper is to present the findings from a systematic review of evaluation studies for physiologic monitoring displays, centered on empirical assessments across all available settings and samples. The findings from this review give readers the opportunity to examine past work across studies and set the stage for the design and conduct of future evaluations.

A broad literature search of the literature from 1991 to June 2007 on PubMed and PsycINFO databases was completed to locate data-based articles for physiologic monitoring device display evaluations. The results of this search plus several unpublished works yielded 23 publications and 31 studies.

Participants were faster detecting an adverse event, making a diagnosis or a clinical decision in 18 of 31 studies. They showed improved accuracy in a clinical decision or diagnosis in 13 of 19 studies and they perceived a decreased mental workload in 3 of 8 studies. Eighteen studies used a within subjects design (mean sample size 16.5), and 9 studies used a between group design (mean group size 7.6). Study settings were usability laboratories for 15 studies and patient simulation laboratories for 6 studies. Study participants were anesthesiologists or anesthesiology residents for 19 studies and nurses for 5 studies.

The advent of integrated graphical displays ushered a new era into physiological

*With kind permission from Springer Science+Business Media: Görges M, Staggers N. Evaluations of physiological monitoring displays: a systematic review. J Clin Monit Comput. 2008;22(1):45-66. ©Springer 2007

monitoring display designs. All but one study reported significant differences between traditional, numerical displays and novel displays; yet we know little about which graphical displays are optimal and why particular designs work. Future authors should use a theoretical model or framework to guide the study design, focus on other clinical study participants besides anesthesiologists, employ additional research methods and use more realistic and complex tasks and settings to increase external validity.

## 2.2 Introduction

The use of physiological monitoring displays is an essential part of clinical care in contemporary health settings. More to the point, the design and interpretation of these displays allows clinicians to detect critical events in a time-sensitive manner, optimally leading to improved patient outcomes. Empirical evaluations of physiological display designs have been published since the early 1990s when computer technology was advanced enough for graphical, real-time monitoring to occur. Yet, no systematic review of the field is currently available.

Two previous, less formal reviews are published. Sanderson et al.[1] discussed advantages and disadvantages of advanced display technology, comparing these display methods for anesthesiology: Advanced visual displays, head-mounted displays, auditory displays and combinations thereof. As part of a literature review of 9 citations through the year 2002, Drews and Westenskow[2] examined previous work on traditional and graphical displays for detection, diagnosis and treatment modalities in anesthesia. Both of these excellent reviews center on anesthesiology. However, nurses are the largest group of clinical display users in clinical settings. This review improves upon previous work by broadening the assessments to all evaluations in all settings, including citations through mid-2007, and employing formal systematic review techniques to analyze past work.

The purpose of this paper is to present the findings from a systematic review of evaluation studies for physiologic monitoring displays, centered on empirical assessments across all available settings and samples. The findings will give readers the opportunity to examine past work across studies and set the stage for the design and

conduct of future evaluations.

## 2.3 Background

The first recording of a human electrocardiogram (ECG) in 1887 and its improvements by Einthoven led to the development of cardiac patient monitors. Computerized ECG was one of the first applications for continuous patient monitoring.[3] Since then, standard cardiovascular patient monitoring has changed little. Only small enhancements, such as color displays or trending (both tabular and graphical) have been incorporated into displays available in the marketplace. A more significant but rather hidden improvement occurred with better alarm algorithms, e.g., outlined by Imhoff and Kuhls,[4] and sensors to reduce the number of false alarms.

Current physiological patient monitoring displays follow the single-sensor, single indicator paradigm, showing one waveform and/or numeric for each sensor.[5] Some sensors provide more than one indicator, such as pulse oximeters or pulmonary artery catheters. Most important, all available monitors still require health care providers to integrate multiple sources of pertinent information in their heads to make an appropriate clinical decision.

Some novel graphical displays are available commercially; however, few have been formally evaluated. Conversely, recent empirical evaluations for proposed integrated displays have been completed, but only two are commercially available in the marketplace currently: (a) an anesthesia drug display evaluated by Syroid et al.[6] and Drews et al.[7] is in the GE CareStation's Navigator Applications Suite (GE Healthcare, Waukesha, WI), and (b) a variation of George Blike's display is in Dräger's Zeus anesthesia workstation (Dräger Medical AG, Germany). The numeric, polygon and histogram displays evaluated by Gurushanthaiah et al.[8] were initially in the Ohmeda Modulus CD anesthesia machine (Ohmeda, Madison, WI now GE Healthcare). However, this anesthesia machine is no longer available and newer versions do not include the novel display. Thus, only two integrated displays in the commercial market have had the benefit of an empirical evaluation.

## 2.4 Methods

A broad literature search of the literature from 1991 to June 2007 was undertaken to locate articles dealing with evaluations of physiologic monitoring device displays. The search began with the year 1991 because the technical capabilities for displays were not advanced enough before then to provide graphical displays. The search was performed on PubMed and PsycINFO databases using the terms found in Appendix A. The search yielded 1,012 (999 on PubMed and 13 on PsycINFO) references. Both authors independently assessed citations for relevancy using the following criteria: (a) physiological monitoring display evaluation, (b) empirical assessment, and (c) English language. Exclusion criteria were: (a) editorials or opinion pieces, (b) descriptions of usage or adoption only, (c) design explanations with no evaluation, (d) review articles, and (e) qualitative research. The raters compared relevancy results and discussed any differences in findings. Where differences existed, the citation was included for further evaluation. Additionally, if relevancy could not be determined from the title, the citation was included in the next step of the relevancy assessment.

From these initial references, 93 articles were identified as being potentially relevant. The authors independently evaluated the abstracts and categorized them into one of the following: relevant, questionably relevant and not relevant. The raters compared the results for agreement; for any discrepancies, the raters discussed each abstract. If any question about relevancy remained, the article was rated as questionably relevant and the full article was retrieved for evaluation. At the end of this process, all articles rated as relevant or questionably relevant were retrieved for further evaluation.

A total of 59 articles were retrieved, read, rated and discussed by the two raters. The articles were rated for relevancy in a dichotomous manner, yielding 18 articles. One additional article,[11] published in late 2007 while this manuscript was under review, was added to the set because of its pertinence. Fugitive literature was included when it was discovered: (a) 2 posters, (b) 1 doctoral dissertation and (c) one 1 paper from a journal (Cognition, Technology & Work) not listed in PubMed or PsycINFO. The final set consisted of 23 references.

## 2.5 Results

The 23 articles matching the relevance criteria are listed in Table 2.1.

Several of the articles reported results of multiple studies; therefore, the total number of completed studies is 31. Each of the studies was evaluated using a quality assessment called QUASII.[29] This new instrument was developed as a tool specifically for assessing empirical studies in clinical informatics. Items are organized around the four ''threats to validity model'' of Cook and Campbell[30] and Shadish, Cook and Campbell[31] and were adapted from the general meta-analytic literature and accepted texts on evaluating research quality.[32–34] During the item development for the instrument, clarification was achieved iteratively, until an inter-rater reliability with a final overall kappa between two raters of 0.85–0.94 was obtained. The QUASII scores for the articles ranged between 78 and 123 out of possible total of 126.

### 2.5.1 Study Settings

Studies were completed in laboratories in Australia, Canada, Germany, Sweden, the United Kingdom and the United States; 12 of 31 were performed at the University of Utah. The most common study settings were usability laboratories (15 studies) or a patient simulation laboratory (6 studies). Two studies were conducted in a naturalistic environment, one on a medical intensive care unit and one in a meeting room of a neonatal intensive care unit. The remaining 8 studies used static computer screens, computer simulations and in 2 cases, paper mock-ups of designs where the setting was immaterial.

### 2.5.2 Study Participants

Researchers used both clinical and nonclinical participants. Nineteen studies used anesthesiologists and anesthesiology residents. Six studies had various nurse, respiratory therapist and/or physician participants. Six study samples were nonclinical—2 each with engineering students, general public and anesthesia staff, and psychology undergraduates.

**Table 2.1**: Physiological monitoring display evaluations

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Agutter et al. (2006)[9] | 30 nurses (15 student nurses and 15 nurses)<br>Static computer screens in a laboratory setting | Design: Within subjects comparing a graphical visualization for arterial blood gas and respiratory values to a traditional numeric display. Nurse expertise as a between groups variable<br>Task: 22 questions about acid–base and respiratory parameters | Time to diagnosis, accuracy, and perceived workload | Faster in responding accurately<br>More accurate in the diagnosis and trending of acid–base questions<br>More accurate in the diagnosis of oxygenrelated parameters<br>Reduced perceived workload | 115<br>Iterative design with usability evaluations of each design<br>Fixed order of events, but one group started with visual graphic and the other with the traditional display |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Agutter et al. (2003)[10] | 20 anesthesiologists<br>Human patient simulator in a simulated operating room | Design: Between groups comparing cardiovascular values on a numeric and a graphical display<br>Task: Two scenarios (anaphylaxis or AP during a total hip replacement and myocardial infarction or MI during a radical prostatectomy); each lasted 10 min, talk-aloud protocol | Times: To detect an adverse event, to diagnosis, to treatment, vital sign deviations and perception of workload | Faster MI detection time with the graphical display but no difference for AP<br>No difference in time to diagnose<br>Faster treatment time for MI using the graphical display<br>Less BP and CVP deviation in MI using the graphical display<br>Users rated the graphical display more useful than the control group<br>No differences in perceived workload | 88<br>Small sample size per cell<br>No assessment of group equivalency<br>Randomized scenario order and display condition<br>Short 10-min scenario |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Albert et al. (2007)[11] | 16 anesthesiologists (7 attendings, three 2nd-year and six 3rd-year residents)<br>Human patient simulator in a simulated operating room | Design: Between groups comparing cardiovascular values on a numeric and a graphical display<br>Task: Five scenarios (mild pain, myocardial ischemia/ infarction or MI, left ventricular failure or LVF, hypovolemia and acute respiratory distress syndrome or ARDS) each lasting 5–9 min, talk-aloud protocol | Expert ranking of performance, times to diagnose and treatment, perception of workload | Improved performance with the graphical display for mild pain, MI and LVF.<br>No difference for hypovolemia and ARDS<br>Faster detection time for MI, LVF and high pulmonary wedge pressure with the graphical display<br>Faster treatment time for MI with the graphical display<br>No effect on perceived workload | 104<br>Small sample size per cell<br>Randomized, counterbalanced design<br>Data from the sepsis scenario was discarded, disrupting the counterbalanced design<br>Short, 5–9-min scenarios |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Blike et al. (2000)[12] | 7 anesthesiologists (5 senior residents and 2 attendings)<br>Static computer screens in a laboratory setting | Design: Within subjects comparing 3 display formats (numeric, object or OD and object minus shapes or OMS)<br>Task: 2 diagnostic tasks in 10 randomly presented scenarios (5 with and 5 without shock) during 2 sessions (Displaysnumeric and OMS, then OD and OMS). | Time to detect shock and accuracy of possible etiology. | Faster detection time with OMS<br>Worse accuracy in recognizing the clinical state with OD<br>Faster etiology determination with the OD<br>Both numeric and OD had higher error rates for etiology determination than OMS | 86<br>Possible order effect due to display.<br>OMS tested twice and etiology time significantly faster in session 2<br>Learning effect as detection time averaged 1.8 in 2nd session versus 2.2 in the 1st one<br>Random order of scenario and display |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Blike et al. (1999)[13] | 11 anesthesiologists (senior residents and attendings)<br>Static computer screens in a laboratory setting | Design: Within subjects comparing graphical object and numeric displays<br>Task: 10 clinical scenarios (5 with and 5 without shock) in a fixed presentation order during separate testing sessions. | Time to decision and diagnostic accuracy of shock/no shock condition. | Faster time to recognize noshock and determine shock etiology with object display<br>Improved diagnostic accuracy with object display<br>Lower proportion of erroneous diagnostic decisions with object display | 106<br>Task simplicity (stated by the author)<br>Could have assessed performance equivalency for levels of physicians<br>Random order for scenarios, fixed display order<br>Potential learning effect as same scenarios were repeated |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Cole and Steward (1994)[14] | 8 respiratory therapists (4 supervisors) using paper sheets | Design: Within subjects comparing a paper graphical metaphor to a table of respiratory values<br>Task: 32 trials judging the patient's respiratory state (4 different states 4 trials x 2 displays). Ordered 2 different ways. | Time to decision and accuracy | Anecdotal report that learning times for the metaphor took less than 5 min<br>Time halved to make a decision with metaphor<br>Similar error rates with both | 94<br>Counterbalanced blocks (4) of 8 trials.<br>Random assignment of subjects to blocks<br>Potential learning effect (only 2 sequences versus random order)<br>Less than 10 min training time for all subjects |
| Doig (2006) [15, study 2] | 30 critical care nurses<br>Static computer screens in a laboratory setting | Design: Between groups comparing a new visual graphic with the standard numeric display<br>Task: 25 multiple response questions based on patient scenarios. Usability questionnaire | Time and accuracy of diagnosis or clinical decision, display usability | No improvement or reduction in data interpretation accuracy<br>Improvements in response accuracy for 2 scenarios, one for each display type<br>Graphical display was favorably rated in terms of acceptance and usability | 94<br>Randomized order of scenarios<br>5–7 min short display training provided for both groups<br>Group equivalency assessed |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Drews et al. (2006)[7] | 30 anesthesiologists with three levels of expertise<br>Human patient simulator in a simulated operating room | Design: Between groups comparing a visual display of real-time drug concentrations to a control group without the display<br>Task: Intravenous anesthesia for simulated shoulder surgery. Surgical plan altered once to increase task complexity | Hemodynamic control of a simulated patient (deviation from baseline vital signs), patient induction, wake up, overall procedure times, perceived workload, satisfaction and subjective utility of the drug display | Significantly less heart rate and blood pressure deviations using drug display<br>2-min faster wake-up time<br>Shorter total procedure times<br>Higher subjective performance with the display<br>No interaction effects for expertise and asks | 118<br>Standardized training for both groups<br>Surgeon interacting with anesthesiologist following prescripted comments, questions and visual cues |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Effken et al. (1997)[16] | Study 1. 18 psychology undergraduates<br><br>Computer simulator in a laboratory setting | Design – Study 1: Between groups comparing 3 displays (traditional strip-chart or TSC, integrated balloon or IBD, and etiological potentials or EPD) showing cardiovascular values<br><br>Task: Three scenarios (low heart strengths, high resistance, low fluid) twice each | Study 1: Time to initiate treatment, number of drugs used, percentage of time in the target range | Study 1: No differences for time to treat<br>Fewer drugs and more time in target vital sign range with EPD<br>Low heart strength scenario showed the greatest time in the vital sign target range | Study 1: 78<br>Psychology students not familiar with clinical tasks<br>Small sample size<br>Training for 20–30 min on each display<br>Use of simulated drugs influencing only 1 parameter each |
| Effken et al. (1997)[16] | Study 2: 11 psychology undergraduates<br><br>Computer simulator in a laboratory setting | Study 2: Same as study 1 using a within subjects design | Study 2: Same as above | Study 2: Faster times to initiate treatment for both IBD and EPD<br>Fewer drugs with EPD overall<br>Fewer drugs with EPD in low fluid scenario<br>Low heart strength scenario showed drugs TSC >IBD > EPD | Study 2: 96<br>Use of psychology students<br>Counterbalanced scenario presentation order, but same display order<br>Training on all displays |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Effken et al. (1997)[16] | Study 3: 6 experienced critical care nurses and 6 nursing students<br>Computer simulator in a laboratory setting | Study 3: Same as Study 2 adding skill levels as a between groups variable. | Study 3: Same as above | Study 3: Faster time to initiate for IBD and EPD with no difference between skill levels<br>Fewer drugs with EPD but fewer drugs in low fluid and heart strength scenarios only<br>Greater time in cardiovascular target with EPD<br>Novices equaled experts' target time performance with IBD<br>No difference in low fluid for IBD and EPD<br>More time in target with EPD in than with the two other displays | Study 3: 109 |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Görges et al. (2006)[17] | 12 2nd- and 3rd-year anesthesia residents using static computer screens<br>Poster presentation | Design: Within subjects comparing three different trend windows (control, simple trend and complex trend)<br>Task: 6 scenarios (control, bronchospasm, pulmonary edema, pneumothorax, pulmonary embolism, malignant hyperthermia, control scenario) | Time to correct diagnosis and perceived workload | No differences in time with a trend toward decreased times for correct diagnosis using simple trend and complex trend | 113<br>Randomized order of events and displays<br>Small sample size<br>Should reanalyze data using repeated measures ANOVA versus Fisher's ANOVA |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Gurushanthaiah et al. (1995)[8] | Study 1: 13 anesthesiology residents (1–4th year)<br>Computer simulator in a laboratory setting | Design – study 1: Combined within subjects comparing 3 displays (polygon, histogram or numeric), and between groups for high (9 trials each per display) and low (4) stimuli. Subsequently, frequency data paired to create a within subjects variable<br>Task: 6 anesthesia scenarios with 10 physiologic variables lasting 6 min during 2 separate sessions | Study 1: Time to detect change, accuracy (which variable and the direction of the change) | Study 1: No effect for time on stimulus frequency or accuracy when analyzed as a between groups variable<br>Faster times for all other residents compared to firstyear residents<br>Faster detection time with the histogram or polygon display<br>Increased accuracy (changed variable and direction of change) with histogram and polygon display<br>Correct identification responses occurred more rapidly than incorrect ones and no difference between identification and direction of change | Study 1: 123<br>Pilot work done<br>Training with competency levels verification to determine adequacy<br>Small sample for between groups design<br>Assessed for confounders (caffeine, alcohol, sleep)<br>Change detection without interpretation of cause |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Gurushan-thaiah et al. (1995)[8] | Study 2: 5 of the same subjects studied in 4 additional sessions | Study 2: Same task, design, displays with additional trials. Randomized, blinded, Latin-squared within groups design with high/low frequency randomized in pairs. | Study 2: Same | Study 2: Faster response time and accuracy for histogram and polygon displays<br>No performance (time or accuracy ) improvement with additional sessions (users were sufficiently practiced) | Study 2: 123<br>Randomized, blinded, crossover, Latin-Square design |
| Gurushan-thaiah et al. (1995)[8] | Study 3: 5 nonmedical volunteers (anesthesia staff) | Study 3: Between groups (anesthesiology users and nonmedical users) | Study 3: Same | Study 3: No differences for time with displays for nonmedical volunteers<br>Decreased accuracy between nonmedical and anesthesia residents with all displays | Study 3: 102 |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Jungk et al. (2000)[18] | Study 1: 16 anesthesiologists<br><br>Anesthesia computer simulator in a usability laboratory | Design-Study 1: Within subjects comparing a simulator monitor with the same monitor plus an ecological interface (EI)<br>Task: Two critical incidents (blood loss and cuff leakage) during a simulated inguinal hernia repair. Eye-tracking and think-aloud protocol. | Study 1: Number of successful trials (identifying critical events), time to identify events; time and frequency of eye fixation on various display regions | Study 1: 43% of the surgery time spent on the EI<br>Faster identification of cuff leakage with EI<br>Equivalent time to identify blood loss in both<br>3 of 8 subjects using the EI missed the blood loss event; none did with the control<br>Eye fixation was diverse | Study 1: 111<br>3 subjects had experience with the EI<br>45min training, familiarization times |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Jungk et al. (2000)[18] | Study 2: 8 anesthesiol-ogists<br>Anesthesia computer simulator in a usability laboratory | Study 2: Within subjects design and same tasks as Study 1 except the use of a redesigned ecological interface display (EI) | Study 2: Time to identify critical events and number of successful trials | Study 2:All correctly identified blood loss but 1 of 8 missed the cuff leakage event<br>Faster identification of both events with the EI | Study 2: 113<br>45 min training or familiarization times<br>All subjects (same subjects as study 1) used the new design. Results compared to the previous study |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Jungk et al. (1999) [19] | 20 anesthesiologists (experts and novices)<br>Static computer screens in a laboratory setting | Design: Within subjects comparing 2 new displays (profilogram or PD and ecological display or ED) to a traditional trend display (TD)<br>Task: Normalizing vital signs from a pathological start state by adjusting sliders. Think-aloud protocol and eye-tracking used. | Ideal circulatory performance (fewer frequency of slider actions, eye tracking parameters, vital sign parameters, and time to completion) | ED accuracy highest. Goal not achieved in 37% of tasks with TD, 19% with PD and 13% with ED.<br>No effect of experience or age on analysis parameters<br>Faster trial time, lower frequency of slider actions and eye fixations for the traditional TD.<br>Correlation between time and entropy (strategic scan paths = system understanding) for ED and TD | 83<br>Unclear whether displays and tasks were counterbalanced<br>Potentially subjects still learning the task with only 2 tasks<br>Analyzed differences between trial 1 & 2<br>Control task not clinically relevant<br>20-30 min training |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Law et al. (2005) [20] | 40 neonatal intensive care unit volunteers (3 levels of nurses and 2 levels of physicians)<br>Static computer screens tested in a meeting room | Design: Within subjects, counter-balanced comparing text summaries to trend graphs for NICU patients<br>Task: 8 medical scenarios each for 2 conditions. Actions selected from a standard list of 18 items. Conditions completed on days 0-31, most in 3-21 days. | Scenario completion time, main expected actions, proportion of correct actions, proportion of nurse and doctor actions, total number of actions and of these the number of appropriate actions | Higher accuracy with text for main actions, proportion of correct ones, nurse/doctor actions, total number of actions and proportion of chosen actions that were appropriate<br>Higher subjective preference for the graphical display<br>No differences in speed of responses, groups or an interaction effect | 112<br>Scenarios may not be equivalent<br>Subjects may remembers scenarios during short intervals<br>No randomized order of events or presentation condition<br>Trends contained information not available in the text presentation |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Liu and Osvalder (2004)[21] | 20 nursing students<br>Static computer screens in a laboratory setting | Design: Within subjects comparing a circular graphical design and numerical reference data<br>Tasks: Six scenarios showing before and after state of a ventilator deviation. Randomized task sequences during 2 testing sessions. | Objective: Change detection time, 3 types of errors (number of deviations, their meaning and the overall situation)<br>Subjective: Deviation severity, reasons for their decision and state opinions about the circular display design. | No differences in detection time.<br>Fewer errors in interpreting the meaning of changes<br>No difference in the number of detected deviations or assessing the overall situation<br>Most preferred and found it easier to detect changes and assess the overall situation with the circular, graphical display | 108<br>Nursing students were new to ventilator issues (construct validity issue)<br>Used a pilot study to optimize study methods<br>Discussed prototype with investigator with added scenarios |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Michels et al. (1997) [22] | 10 anesthesiologists<br><br>Anesthesia computer simulator in a laboratory setting | Design: Between groups comparing graphical to traditional numeric and waveform display of physiological variables<br>Task: 4 critical events (blood loss, inadequate paralysis, endotracheal tube cuff leak, depletion of soda lime) | Detection time and correct identification of critical anesthesia events | Results dependent upon clinical event<br>Faster detection for 2 of 4 events (inadequate paralysis and cuff leak) with graphical display<br>Correct identification sooner for 3 of 4 events (paralysis, cuff leak and blood loss) with graphical display | 94<br>Very small sample per cell (5)<br>No assessment for group equivalency<br>Same sequence of scenarios used for each participant<br>15 min introduction to displays<br>Alarms silenced to rely on visual observations only |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Ng et al. (2005)[23] | 10 engineering students<br>Simulated clinical setting in a usability laboratory | Design: Within subjects comparing 3 alarms: vibro-tactile, auditory alarm and a combination of the two.<br>Task: 24 randomly generated alarm events for training. 30 events during a 30 min interval based on real clinical data using 6 simulated alarm patterns in three levels of severity. Subjects trained to recognize 6 alarm patterns | Training, identification rate (number of events detected), accuracy of alarm patterns, response time, comfort and satisfaction | No difference in number of training alarms required to learn display alarms<br>Higher identification rate with the vibro-tactile than audible or combined alarm display<br>Higher identification rate for combined than auditory alone<br>No difference in time to respond to an alarm<br>Perception that vibro-tactile would attract attention more readily<br>Preference for vibro-tactile (4) than auditory (3) or combination (3)<br>Reduced accuracy for combined than vibro-tactile alone (for Level 1 alarm only)<br>90% of the subjects reported some discomfort with the vibro-tactical alarms.<br>Subjects preferred the vibro-tactile alarm despite the discomfort | 107<br>Use of engineering students performing clinical tasks<br>Auditory accuracy for level 1 alarm only<br>Pilot study used to optimize vibro-tactile display<br>Randomized display order. Under if scenarios randomized |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Syroid et al. (2002)[6] | 15 anesthesiologists (seven attendings, three 2nd-year and five 3rd-year residents)<br>Anesthesia computer simulator in a laboratory setting | Design: Within subjects, counter-balanced with and without a graphic display showing intravenous drug concentrations<br>Tasks: 2 clinical scenarios (abscess drainage and mass removal) using the same 3 drugs | Precision in drug administration, number of bolus doses, vital signs to indicate pain response, and perceived workload. | Lower variation (tighter control) in the effect-site concentrations of anesthetics with the drug display<br>During maintenance, more remifentanil doses given with the drug display<br>No differences in propofol boluses<br>No differences in vital signs (pain levels)<br>Perceived decreased mental demand, frustration, effort and increased performance with the drug display | 116<br>Subjects commented that the bolusing of anesthetic agents was not realistic<br>Randomized scenario and display order<br>Simulation required extra effort to obtain patient responses<br>Low task complexity, short scenarios, artificial simulation |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Wachter et al. (2006)[24] | 19 clinical volunteers (nine anesthesia faculty, four 2nd-year residents and six 3rd-year residents) from 2 universities Patient simulator in a usability lab | Design: Between groups comparing a pulmonary graphical display to traditional numeric displays<br>Task: Five scenarios (4 adverse, obstructed endotracheal tube, endobronchial intubation, intrinsic PEEP, hypoventilation; 1 normal event). | Time to correct diagnosis, time to treatment (experts viewed videotapes) and perception of workload | Faster detection and treatment times for 2 of 4 events - obstructed endotracheal tube and intrinsic PEEP events using the graphical display.<br>Unnecessary treatment given by 3 clinicians using the graphical and 5 using numerical display<br>No difference in diagnostic accuracy<br>Lower subjective workload for obstructed endotracheal tube and intrinsic PEEP scenarios. | 89<br>No assessment of group equivalency. Did not measure critical individual differences<br>Pilot study used to determine adequate training time<br>Randomized order of events<br>No data about group equivalency<br>No discussion about unnecessary treatments |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Wachter et al. (2005)[25] | 32 caregivers (critical care physicians, nurses and respiratory therapists)<br>Pulmonary metaphor graphical display used in an actual intensive care unit | Design: Descriptive<br>11 day observational study of display use in a medical intensive care unit. | Display observations per caregiver visit, perceived usefulness, acceptance, desirability and accuracy of the display | Profession/number of times entering the room/number of display observations per visit<br>Nurses/ 775/ 1.3,<br>Respiratory therapists (RTs)/ 74/ 3<br>Physicians/ 34/ 6<br>Physicians and RTs looked at the display more often over the course of the study<br>No difference in questionnaire response for caregiver groups<br>Perceptions ranged from 5-6.5 (0-9 scale on usefulness, desirability, accuracy and acceptance) | N/A, Descriptive Study<br>Display provided new (etCO2) information not available to caregivers beforehand<br>Mid-scale perception ratings interpreted as positive |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Wachter et al. (2003)[26] | 46 clinicians (22 anesthesiologists,1 nurse anesthetists, 18 residents and 5 medical students from 3 facilities)<br>Static computer screens in a laboratory setting | Design: Descriptive for 5 design iterations for a pulmonary graphical display evaluated using paper-based tests | Correct identification of pulmonary design components to anatomical parts and pulmonary variables, ability to diagnose pulmonary events. | Improved anatomical intuitiveness by 25% (to 98%) and variable mapping intuitiveness by 34% (to 91%) for 5th design<br>Fifth design decreased diagnostic accuracy by 4%. (to 79%). | N/A, Descriptive Study<br>Use of multiple choice tests limited choices for subjects<br>Different compositions of iteration testing groups as well as different sample sizes<br>Participants not given waveforms or history for displayed values available |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Watson and Sanderson (2004)[27] | Study 1: 23 paid general public participants (7 men, 16 women)<br>Laboratory setting | Design-Study 1: Within subjects comparing 3 recorded respiratory sonifications for 3 conditions (respiratory rate or RR, end-tidal carbon dioxide or etCO2 and tidal volume or VT).<br>Task: 12 anesthesia scenarios (3 for training) lasting 4.5-5min each with physiological events and mechanical changes | Study 1: Assessing abnormality (high, low or normal value) and direction (increasing, decreasing or steady), confidence of judgment and perception of workload | Study 1:Improved abnormality assessment with the varying sonification, especially for sonification of etCO2 and VT, which also had a slight preference in user preference.<br>No effect for direction judgments<br>Subjects preferred the varying tone for RR, VT and etCO2<br>No workload effect | Study 1: 96<br>Use of the general public for a clinical task<br>Large age range (19-55)<br>Possible order effect<br>Use of prerecorded audio files without scenario randomization |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Watson and Sanderson (2004)[27] | Study 2: 11 anesthesiologists and 10 information technology postgraduates<br><br>Laboratory setting | Design-Study 2: Within subjects, same objectives.<br>Task: Six scenarios with fewer abnormal changes than Study 1 | Study 2: same as in study 1 | Study 2:Improved abnormality judgments and direction for anesthesiologists than IT postgraduates.<br>Anesthesiologists had higher perceived workload but not significantly | Study 2: 104<br>Arithmetic control task<br>Use of prerecorded audio files. No scenario randomization |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Watson and Sanderson (2004) [27] | Study 3: Same participants as in study 2<br>Laboratory setting | Design-Study 3: Same design and objectives<br>Task: Nine scenarios lasting approximately 9 min each, using a computer simulation with sonification alone (S), visual display (V) and combined (SV). Used a distracter task of arithmetic determinations. Added an additional alarm for heart rate or HR. Arithmetic accuracy communicated as the main study goal | Study 3: same as in study 1 | Study 3:Improved abnormality judgment main effect with SV, then V, then S but no effect for anesthesiologists<br>Higher abnormality judgment with HR task and least with VT<br>Anesthesiologists performed better than IT postgrads<br>Less directional accuracy with VT than other events<br>Higher confidence in O2 judgments and lowest in RR<br>Anesthesiologists preferred the combined mode although it was perceived to have the highest workload | Study 3: 104<br>Quasi-randomized query for parameters<br>Potential learning effects |

**Table 2.1** continued

| Source | Sample, setting | Study design, tasks | Dependent variable(s) | Key findings | QUASII score and quality considerations |
|---|---|---|---|---|---|
| Zhang et al. (2002) [28] | Study 1: 12 anesthesiologists (attending and residents)<br>Human patient simulator in a simulated operating room | Design-Study 1: Within subjects comparing Blike's 3-D object display to traditional numerical display<br>Task: 6 scenarios in random order for training. Four 10-min events (hypovolemia, myocardial ischemia, arrhythmia, bronchospasm) | Study 1: Time to recognize event, time to diagnose and situational awareness (SA) scores | Study 1:No difference in event recognition time for cardiovascular events<br>Faster detection times for bronchospasm with the 3-D object display.<br>Interaction effect: Intermediate level SA scores greater for hypovolemia with the object display.<br>Interaction effect: Low level SA scores greater during arrhythmia, hypovolemia and bronchospasm with traditional displays | Study 1: 105<br>Issues with training, practice<br>Potential order effect for displays<br>Randomized scenario order<br>Simulation freeze technique to allow subjects to answer questionnaires<br>Scenarios had different difficulty levels |

Nine of the 31 studies reported the sample's mean age, ranging from 31–42.6 years. In one paper[27] the ages of the nonclinical samples vary from 19–55 and 29–62 in comparison to the clinician group's age range of 23–44 years. Six of the 31 studies report the expertise of participants in mean postgraduate years, ranging from 5–13.9 years. Ten studies did not report expertise while 13 studies include samples with 2 or more levels of expertise. Doig[15] mentioned that study groups were balanced for intensive care nurses' expertise. Other participant variables were measured: 5 studies measured hr of sleep in the previous night, 5 reported participants' caffeine and medication consumption and 1 obtained additional measures such as color vision, vision quality, and dominant hand.

Average sample sizes ranged from 5–46 subjects. Within subjects designs had a mean sample size of 16.5 while between group designs had an average of 7.6 participants per cell. Total sample sizes for between groups studies ranged from 5 to 30.

### 2.5.3 Display Type

A variety of displays were studied: 13 hemodynamic/cardiovascular, 6 pulmonary/ respiratory, 4 integrated anesthesia and 2 anesthesia drug graphical displays, 3 respiratory sonifications, and 1 each vibro-tactile and sonification display, arterial blood gas graphic and physiologic trend graphic. All but Görges et al.[17] reported significant improvements for accuracy and/or speed with the new designs.

### 2.5.4 Study Design

Eighteen studies used a within subjects design while 9 used a between groups design. Two studies employed combined designs (both within subjects and between groups), and two other studies were descriptive (an observation and a description of design iterations for a pulmonary metaphor). Twenty-one studies randomized (or counterbalanced) scenario order and 10 randomized display order. In fact, Gurushanthaiah et al.[8] used Latin-squared randomization to guide the order of tasks.

### 2.5.5 Tasks

Fifteen studies devised anesthesia scenarios and 2 others used medical decision tasks. Seven studies used deviation or event detection tasks while 2 studies used multiple choice questions about respiratory events. The 2 descriptive studies outlined the use of the display in normal clinical workflow.

Nonclinical participants worked with the clinical scenarios in 6 studies. These participants included psychology students,[16] nonmedical anesthesia staff,[8] engineering students,[23] the general public and IT postgraduates,[27] and bioengineering students.[28]

Twenty-two authors reported giving training to participants while 2 studies provided ''instruction.'' Nineteen authors reported that participants were allowed to practice with the new device. The combination of practice and training with displays lasted from 2–45 min. One author allowed more practice if participants did not meet cut scores. Seven authors either used cut scores for admitting participants into the study or had participants practice until specific performance goals were met.

### 2.5.6 Dependent Variables

The most common dependent variable was time to complete a task (make a diagnosis, detect an adverse event or initiate treatment), measured in 30 of the 31 evaluation studies. Participants were faster detecting an adverse event or making a diagnosis or decision in 18 studies.[7–14, 18, 19, 22, 24, 28] Participants in 13 of 19 studies showed improved accuracy in a clinical decision or diagnosis.[8, 9, 11, 13, 15, 19–23, 27] Five studies used a control task, measuring the percentage of time spent within a target range or deviations in vital signs. With graphical designs, participants[6, 7, 10, 16] had less vital sign deviations or deviations from a target range. Three of 8 studies showed decreased perceived workload, with a graphical design,[6, 9, 24] and 3 studies described screen display regions of interest measured with an eye tracker. Other dependent variables included 3 studies measuring satisfaction, subjective utility, situational awareness, display usefulness and whether the scenario was realistic. Overall, these studies demonstrated the positive impacts of a graphical design on speeding clinician time to detect an event, determine a diagnosis, determine a correct diagnosis and stay within a target range of variables.

## 2.6 Discussion

None of the studies reported using a theoretical model or framework to guide the study or its methods although a number of theoretical works are now available.[35–39] Theoretical models or frameworks are organizing structures researchers can use to assist with study design. These conceptual structures allow researchers to consider major variables of interest as well as potential confounding variables. For instance, frameworks with a developmental timeline[37, 38] remind researchers to consider both practice and training because users and technology change over time. Likewise, individual characteristics guide researchers to measure and/or control for participant differences. These kinds of elements might appear straightforward to readers; however, these variables were not consistently reported or considered in published studies.

### 2.6.1 Study Settings

The most common settings for studies were usability laboratories or those simulating operating rooms (ORs). However, practicing clinicians use monitors in a number of settings besides the OR, e.g., emergency departments, telemetry units, intensive care units, and prehospital modes of transportation such as air transport and ambulances. In particular, pediatric units, neonatal displays, and even battlefields are not represented in available studies. Remote monitoring of critical care patients, e.g., as outlined by Breslow et al.,[40] is a relatively new care delivery method, presenting a novel setting for future evaluations. With the exception of select intensive care units, settings mentioned here are as yet unexplored or simulated in usability laboratories.

Drews and Westenskow[2] noted that, at this point, researchers cannot be clear about how the studies performed in lab settings correlate to participants' performance in actual clinical settings. The combination of embedding the participant into a more realistic environment, like a simulated clinical setting with a human patient simulator, is a good step forward; however, researchers will want to test their displays in actual clinical settings as well.

### 2.6.2 Study Participants

Anesthesiologists comprised 61% of the total participants in past studies. Displays are not yet designed and evaluated for the largest group of monitor users: Nurses. Their concerns and tasks are distinct from anesthesiologists, so designs are needed for nurses' particular tasks and mental models. More important, current commercial physiological displays do not supporting a walk-by, at-a-glance assessment of the patient's status, a benefit needed by nurses as they multitask during patient care. Respiratory therapists (RTs) are another group of understudied monitor users.

Display users in various settings will not be homogeneous even within professions. For instance, nurses performing trauma care in the emergency department may require different display designs than nurses in intensive care units with the more routine monitoring that occurs there. Likewise, physicians other than anesthesiologists have not been included in evaluation studies, except in two studies.[20, 25]

Participant demographics and individual characteristics are inconsistently reported and/or controlled.[2] Age was not reported in 18 studies and caffeine intake was not reported in 23 studies. Expanding upon that notion, the age range of study participants, when reported at all, varied as much as 30 years. Factors such as age and caffeine intake may be potential confounding variables in studies using response times as a dependent variable. For example, Gurushanthaiah et al. [8, study 3] reported an influence of age and caffeine consumption on participant response times for nonclinical volunteers. Age and caffeine did not influence their results for clinicians; however, the sample size of 5 was very small. Response time and age are positively correlated so including participants in their 50s or 60s should be carefully considered in the future and a more narrow age range should be contemplated. Expertise is another important variable to track or control, especially if a between-groups experimental design is used. Levels of expertise may be a confounder to the observed results, particularly when students are combined with more seasoned clinicians. Future researchers should routinely report participant demographics and pertinent variables such as caffeine intake.

Last, using nonclinical participants, while convenient, raises questions about the external validity and significance of the results. That anesthesiologists out-performed

IT professionals or the general public is not surprising.

### 2.6.3 Study Designs

The majority of studies used within subjects designs. These are particularly well suited to studies involving response time because they control for individual differences which can vary widely across users. Studies using between groups designs received lower quality ratings primarily due to the control for individual differences and the larger sample size required to assure adequate power. Six of the 9 studies with a between groups design had fewer than 15 participants per cell (mean = 7.6) and did not assess group equivalence. No researcher reported conducting a power analysis. Without a power analysis, researchers should have at least 15 per cell in a between group study to assure adequate power.[41]

### 2.6.4 Tasks and Scenarios

A few authors reported validity assessments for clinical scenarios, e.g., Blike et al.[12] or Doig,[15] using clinical experts to validate scenarios or consulting sample case studies from the medical literature. Other authors shortened scenarios for study purposes, e.g., Syroid et al.[6] or Wachter et al.[24] While these abbreviated scenarios are likely to increase the mental workload, they artificially condense time frames,[2] which may confuse the study participant or cause them to eliminate potentially correct diagnoses. Future researchers can learn from these examples by including a scenario validity assessment, e.g., using external experts and considering the use of more realistic scenarios.

Multiple scenarios are likely to have different levels of complexity, e.g., detecting bronchospasm compared to detecting an arrhythmia[28] or detecting bronchospasm compared to detecting a pulmonary embolism.[17] Differences in task complexity need to be assessed and controlled for carefully, as they may become additional covariates that can mask valid results. Once understood, complexity levels can either be randomized to reduce an order effect or controlled across groups to assure equivalency. Of course, tasks can only be randomized if this technique does not destroy the clinical relevancy of the scenario. Otherwise, several scenarios can be presented with equivalent tasks in differing order.

Low mental workload is common across current studies. Displays were essentially isolated from other stimuli, merely showing waveforms and numeric information of the different sensors familiar to clinicians. In most studies, participants can focus exclusively on the required control or diagnostic task without competing demands. Sanderson et al.[1] warn that new displays reveal higher order properties of patient states, yet their benefits in high mental workload situations is unknown. In a realistic environment, a clinician often takes care of more than one patient and may need to perform several tasks at once. Attention to clinician mental workload is needed in the future.

New designs may include variables not typically measured in the clinical setting, creating a dilemma for designers.[22] Choices are: (a) to not display certain elements of the design, (b) to not show the display at all, or (c) to assume values in order for the display to function, all which might pose substantial problems for obtaining FDA approval. Albert et al.[11] offered one solution: condensing the Agutter et al.[10] display by the missing variables while preserving the overall metaphor.

Future researchers can eliminate nonclinical control tasks such as arithmetic distracter tasks, e.g., as in.[19, 27] These do not assist with the external validity of the study and they create a different mental workload than typical clinical tasks. More relevant control tasks are participants' pagers beeping during the scenario, staff talking to the participant during the task, overhearing staff cell phone conversations and other ambient noise. Interruptions are a common occurrence in all settings, yet only a few studies[7, 10, 24] integrated disruptions and distractions into their simulated or actual study settings, e.g., having an investigator distract and interrupt the participant by acting like a surgeon. Scenarios with distractions and requirements for multitasking[7, 42] provide for more realistic environments for participants and aid in requirements development for designers.

Seven studies used cut-scores to test training adequacy before participants were admitted to the study. Cut-scores or other competency assessments can be useful for future researchers to decrease individual differences and variability across subjects. Pilot tests are particularly useful to test study methods, training requirements and to determine the number of tasks to display to ensure adequate practice. Researchers can

display performance times plotted against tasks to observe the resulting performance curves. When the performance curve flattens, the number of tasks and practice is adequate.

### 2.6.5 Future Display Evaluations

Thirty of 31 studies reported significant findings with the new display. This is likely a publication bias; however, from the collected studies, one might surmise that any novel design is a significant one. The next logical step may be to compare graphical designs to each other to find out why particular designs are significant. Additionally, adding a qualitative portion to a study could identify why users find particular designs optimal. Sanderson[43] cites an interview with Matt Weinger about future patient monitoring that would provide real-time, continuous information on organ functions down to the cellular level. Designers will be challenged to integrate vast numbers of values into logical displays to aid clinical decision-making under time pressure.

The NASA-TLX[44] is a tool used in 6 studies. The tool measures various aspects of perceived mental workload, is easy for participants to use, and provides another dimension to users' work with displays. The development of this instrument is described in an original paper[44] and a comparison with alternative methods of workload assessments instruments can be found in Rubio et al.[45] Future researchers may wish to incorporate one of these tools into their work and also perform formal psychometric testing for the instrument to build upon the fine conceptual development of this tool.

All studies to date have examined only the dyad of user and display. However, clinicians typically work as teams in clinical environments. How a monitor might be devised to address the work of teams has not been studied. Last, the opportunities for future researchers are great because many currently available displays lack empirical evaluations.

## 2.7 Conclusions

The advent of integrated graphical displays ushered a new era into physiological monitoring display designs. This systematic review analyzed 31 studies of these novel designs. All but one study reported significant differences between traditional,

numerical displays and novel displays using graphs or sound – decreasing the time to detect an event or the time to make a diagnosis or increasing the accuracy of the diagnosis. Yet we know little about which graphical displays are optimal and why particular designs work. Most studies focused on anesthesia-related participants while future work can explore nurses, respiratory therapists, nonanesthesia physician users as well as teams of users. The majority of current studies were conducted in laboratory settings. In the future, more realistic, complex tasks and settings would provide greater external validity for studies. Most acute care clinical settings and concomitant tasks in emergency departments, pediatric units, ambulances, neonatal intensive care units, and even battlefields are, as yet, unexplored. Future researchers can improve their studies by: (a) Using a theoretical model or framework to guide the study, (b) Reporting and controlling for individual differences of participants, (c) Completing validity assessments of clinical scenarios to ensure clinical realism, (d) Assuring adequate power in the study by conducting a power analysis to estimate numbers of required participants, and (e) Adding a qualitative component to studies in order to better understand how designs work for clinical decision-making.

## 2.8 PubMed Search Terms

PubMed search terms (‘‘computer simulation’’[MeSH] OR ‘‘data display’’[MeSH] OR ‘‘monitoring, physiologic’’ [MeSH:noexp] OR ‘‘patient Journal of Clinical Monitoring and Computing simulation’’ [MeSH] OR ‘‘user–computer interface’’ [MeSH] OR ‘‘models, biological’’[MeSH:noexp] OR ‘‘computer graphics’’[MeSH])

AND (‘‘blood pressure’’[MeSH] OR ‘‘heart rate’’ [MeSH] OR ‘‘intubation, intratracheal/instrumentation’’ [MeSH] OR ‘‘hemodynamic processes’’[MeSH] OR ‘‘respiration’’[MeSH] OR ‘‘respiration, artificial’’[MeSH] OR ‘‘anesthesiology’’[MeSH] OR ‘‘Anesthetics’’[MeSH] OR ‘‘Critical Care’’[MeSH] OR ‘‘Intensive Care Units’’ [MeSH])

AND (ecological[tiab] OR graphic[tiab] OR graphics [tiab] OR graphical[tiab] OR GUI[tiab] OR visual[tiab] OR simulator[tiab] OR simulation[tiab])

AND English[lang]

AND (‘‘1991/01/01’’[EDAT] : ‘‘2007/06/01’’ [EDAT])

AND ‘‘Journal Article’’[ptyp]

## 2.9 Acknowledgments

The authors would like to thank Dr. Dwayne Westenskow for his thoughtful comments on a previous draft of this manuscript.

## 2.10 References

1. Sanderson PM, Watson MO, Russell WJ. Advanced patient monitoring displays: tools for continuous informing. Anesth Analg. 2005 Jul;101(1):161–168.

2. Drews FA, Westenskow DR. The right picture is worth a thousand numbers: data displays in anesthesia. Hum Factors. 2006 Spring;48(1):59–71.

3. Jenkins JM. Computerized electrocardiography. Crit Rev Bioeng. 1981 Nov;6(4):307–350.

4. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg. 2006 May;102(5):1525–1537.

5. Goodstein L. Discriminative display support for process operators. In: Rasmussen J, Rouse W, editors. Human Detection and Diagnosis of System Failures. Springer; 1981. p. 433–449.

6. Syroid ND, Agutter J, Drews FA, Westenskow DR, Albert RW, Bermudez JC, et al. Development and evaluation of a graphical anesthesia drug display. Anesthesiology. 2002 Mar;96(3):565–575.

7. Drews FA, Syroid N, Agutter J, Strayer DL, Westenskow DR. Drug delivery as control task: improving performance in a common anesthetic task. Hum Factors. 2006 Spring;48(1):85–94.

8. Gurushanthaiah K, Weinger MB, Englund CE. Visual display format affects the ability of anesthesiologists to detect acute physiologic changes. A laboratory study employing a clinical display simulator. Anesthesiology. 1995 Dec;83(6):1184–1193.

9. Agutter J, Albert R, Syroid N, Doig A, Johnson K, Westenskow D. Arterial Blood Gas Visualization for Critical Care Clinicians. In: Proceedings of the Annual Meeting of the Society for Technology in Anesthesiology. San Diego, CA; 2006. .

10. Agutter J, Drews F, Syroid N, Westneskow D, Albert R, Strayer D, et al. Evaluation of graphic cardiovascular display in a high-fidelity simulator. Anesth Analg. 2003 Nov;97(5):1403–1413.

11. Albert RW, Agutter JA, Syroid ND, Johnson KB, Loeb RG, Westenskow DR. A simulation-based evaluation of a graphic cardiovascular display. Anesth Analg. 2007 Nov;105(5):1303–1311.

12. Blike GT, Surgenor SD, Whalen K, Jensen J. Specific elements of a new hemodynamics display improves the performance of anesthesiologists. J Clin Monit Comput. 2000;16(7):485–491.

13. Blike GT, Surgenor SD, Whalen K. A graphical object display improves anesthesiologists' performance on a simulated diagnostic task. J Clin Monit Comput. 1999 Jan;15(1):37–44.

14. Cole WG, Stewart JG. Human performance evaluation of a metaphor graphic display for respiratory data. Methods Inf Med. 1994 Oct;33(4):390–396.

15. Doig AK. Graphical Cardiovascular Display for Hemodynamic Monitoring [PhD Thesis]. University of Utah. Salt Lake City, UT; 2006.

16. Effken JA, Kim NG, Shaw RE. Making the constraints visible: testing the ecological approach to interface design. Ergonomics. 1997 Jan;40(1):1–27.

17. Görges M, Förger K, Westenskow DR. Trend Based Decision Support System For Anesthesiologists Improves Diagnosis Speed and Accuracy. In: Proceedings of the Annual Mountain West Biomedical Engineering Conference. Snowbird, UT; 2006. .

18. Jungk A, Thull B, Hoeft A, Rau G. Evaluation of two new ecological interface approaches for the anesthesia workplace. J Clin Monit Comput. 2000;16(4):243–258.

19. Jungk A, Thull B, Hoeft A, Rau G. Ergonomic evaluation of an ecological interface and a profilogram display for hemodynamic monitoring. J Clin Monit Comput. 1999 Dec;15(7-8):469–479.

20. Law AS, Freer Y, Hunter J, Logie RH, McIntosh N, Quinn J. A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. J Clin Monit Comput. 2005 Jun;19(3):183–194.

21. Liu Y, Osvalder AL. Usability evaluation of a GUI prototype for a ventilator machine. J Clin Monit Comput. 2004 Dec;18(5-6):365–372.

22. Michels P, Gravenstein D, Westenskow DR. An integrated graphic data display improves detection and identification of critical events during anesthesia. J Clin Monit. 1997 Jul;13(4):249–259.

23. Ng JYC, Man JCF, Fels S, Dumont G, Ansermino JM. An evaluation of a vibro-tactile display prototype for physiological monitoring. Anesth Analg. 2005 Dec;101(6):1719–1724.

24. Wachter SB, Johnson K, Albert R, Syroid N, Drews F, Westenskow D. The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. J Am Med Inform Assoc. 2006 Nov;13(6):635–642.

25. Wachter SB, Markewitz B, Rose R, Westenskow D. Evaluation of a pulmonary graphical display in the medical intensive care unit: an observational study. J Biomed Inform. 2005 Jun;38(3):239–243.

26. Wachter SB, Agutter J, Syroid N, Drews F, Weinger MB, Westenskow D. The employment of an iterative design process to develop a pulmonary graphical display. J Am Med Inform Assoc. 2003 Jul;10(4):363–372.

27. Watson M, Sanderson P. Sonification supports eyes-free respiratory monitoring and task time-sharing. Hum Factors. 2004 Fall;46(3):497–517.

28. Zhang Y, Drews FA, Westenskow DR, Foresti S, Agutter J, Bermudez JC, et al. Effects of Integrated Graphical Displays on Situation Awareness in Anaesthesiology. Cognition, Technology & Work. 2002 Jun;4(2):82–90.

29. Phansalkar S, Staggers N, Weir C. Development of the QUASII (QUality Assessment of Studies in Informatics Implementations) Instrument. In: VA HSR&D National Meeting. Washington, DC; 2006. .

30. Cook TD, Campbell DT. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin; 1979.

31. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin; 2002.

32. Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994.

33. The Cochrane Collaboration. The Cochrane Manual; 2007 [updated 8/23/2007; cited 9/18/2007]. Available from: http://www.cochrane.org/admin/manual.htm.

34. Shadish WR, Fuller S, editors. The Social Psychology of Science. New York: Guilford Press; 1994.

35. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak. 2006;6:3.

36. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care. 2006 Dec;15 Suppl 1:50–58.

37. Despont-Gros C, Mueller H, Lovis C. Evaluating user interactions with clinical information systems: a model based on human-computer interaction models. J Biomed Inform. 2005 Jun;38(3):244–255.

38. Staggers N. Human-computer interaction. In: Englebardt S, Nelson R, editors. Information Technology in Health Care: An Interdisciplinary Approach. Harcourt Health Science Company; 2001. p. 321–345.

39. Daniels J, Fels S, Kushniruk A, Lim J, Ansermino JM. A framework for evaluating usability of clinical monitoring technology. J Clin Monit Comput. 2007 Oct;21(5):323–330.

40. Breslow MJ, Rosenfeld BA, Doerfler M, Burke G, Yates G, Stone DJ, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: an alternative paradigm for intensivist staffing. Crit Care Med. 2004 Jan;32(1):31–38.

41. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. Boston, MA: Houghton Mifflin; 2003.

42. Strayer DL, Drews FA, Crouch DJ. A comparison of the cell phone driver and the drunk driver. Hum Factors. 2006 Summer;48(2):381–391.

43. Sanderson P. The multimodal world of medical monitoring displays. Appl Ergon. 2006 Jul;37(4):501–512.

44. Hart S, Staveland L. Development of NASA-TLX (Task Load Index) Results of Empirical and Theoretical Research. In: Hancock P, Meshkati N, editors. Human Mental Workload. Amsterdam: North Holland Press; 1988. p. 139–183.

45. Rubio S, Diaz E, Martin J, Puente J. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. Applied Psychology: An International Review. 2004;53(1):61–86.

# CHAPTER 3

# IMPROVING ALARM PERFORMANCE IN THE MEDICAL INTENSIVE CARE UNIT USING DELAYS AND CLINICAL CONTEXT*

## 3.1 Abstract

In an intensive care unit, alarms are used to call attention to a patient, to alert a change in the patient's physiology, or to warn of a failure in a medical device; however, up to 94% of the alarms are false. Our purpose in this study was to identify a means of reducing the number of false alarms.

An observer recorded time-stamped information of alarms and the presence of health care team members in the patient room; each alarm response was classified as effective (action taken within 5 min), ineffective (no response to the alarm), and ignored (alarm consciously ignored or actively silenced).

During the 200-hr study period, 1271 separate entries by an individual to the room being observed were recorded, 1214 alarms occurred and 2344 tasks were performed. On average, alarms occurred 6.07 times per hr and were active for 3.28 min per hr; 23% were effective, 36% were ineffective, and 41% were ignored. The median alarm duration was 17 sec. A 14 sec delay before alarm presentation would remove 50% of the ignored and ineffective alarms, and a 19 sec delay would remove 67%. Suctioning, washing, repositioning, and oral care caused 152 ignored or ineffective ventilator alarms.

---

*With kind permission from Wolters Kluwer Health / Lippincott, Williams & Wilkins: Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg. 2009 May;108(5):1546-52. 

Introducing a 19 sec alarm delay and automatically detecting suctioning, repositioning, oral care, and washing could reduce the number of ineffective and ignored alarms from 934 to 274. More reliable alarms could elicit more timely response, reduce workload, reduce noise pollution, and potentially improve patient safety.

## 3.2 Introduction

Intensive care unit (ICU) alarms were designed to call attention to a patient, to alert a change in the patient's physiology or to alert staff to a device problem. Alarms are triggered when a physiologic variable crosses a set threshold. In their excellent literature review, Imhoff and Kuhls report alarm frequencies of 1.6 to 14.6 alarms/hr and a false alarm rate of up to 90%.[1] Chambrin et al.[2] reported the lowest rate of alarms at 1.6 alarms/hr; however, their study did not include infusion pumps (InfP) or alerts. Tsien and Fackler[3] reported one of the highest alarm rates at 9.8 alarms/hr in a noisier environment, but limited their study to alarms from the cardiac patient monitor. The problem with simple threshold alarms is that up to 94.5% of the alarms that sound in the ICU are false, are provider-induced,[4] and frequently sound unnecessarily.[1, 2, 4] Default settings by the equipment manufacturers are set to avoid missing a single false negative alarm and thereby result in many false positive alarms.[5]

New alarm algorithms and improvements in sensors are reported to reduce the number of false alarms, but many of these suggestions have not been incorporated into current monitors nor have their improvements been evaluated in patients.[1] Rheineck-Leyssius and Kalkman[6] proposed a highly effective method for reducing pulse oximeter (Spo2) alarms by introducing a 6 sec delay thereby reducing alarm rates by 50%. One of the new and interesting approaches to reducing the number of false alarms is the use of context awareness.[7, 8] Dey[8] defines context-awareness as: "A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task." Chambrin et al.[2] report that 42% of the transient ICU alarms are triggered by patient movement or respiratory effort. Therefore, an alarm system that knows the patient is moving or coughing could suppress many motion induced alarms. Although other investigators[2–4, 9, 10]

have classified false alarms into general categories, such as "staff manipulation" or "the patient," we propose using specific tasks performed by the health care provider and each patient's current condition and actions. Some work regarding alarms and their context has been performed. For example, Seagull and Sanderson[11] investigated anesthesia alarms in the context of the surgical phase (induction, maintenance, emergence). However, there is still more to explore in the ICU setting.

The purpose of this study was to observe alarms in the medical ICU (MICU) to identify methods for reducing the number of false alarms by using time delays and the correlations between alarms and clinical context.

## 3.3 Methods

Approval was obtained from the University of Utah Health Sciences Center's IRB and informed consent was obtained from 22 participating health care team members.

At the beginning of each day, for 24 days, the investigator randomly selected a patient room in the MICU, where a tracheally intubated patient was receiving respiratory support. A different patient and room were chosen every morning, except one patient who was observed twice. The investigator recorded health care team members' actions while they were in the patient's room and whether they came into the room in response to an alarm. Health care team members included attending physicians, fellow physicians, resident physicians, nurses, respiratory therapists, health care assistants, physical therapists, medical students, pharmacists, and other providers. Observations began at approximately 7:30 am and ended before 7 pm.

### 3.3.1 Setting

The 12-bed adult MICU is organized in an H shape, with individual patient rooms to the north and south, a central station in its center, and additional function rooms between the two rows of rooms. The doors to the patient's rooms were left open unless procedures were performed or privacy was required. The unit was staffed with one nurse for every two patients, one health care assistant, and one health unit coordinator. Respiratory therapists checked a patient's ventilator when paged or at least once every 4 hr. Most patients had sepsis, respiratory failure, acute respiratory distress syndrome, multisystem organ failure, or renal failure. Approximately 25% of

the patients had myocardial infarction, cardiomyopathy, or arrhythmias.

A cardiac monitor with at least electrocardiography, Spo2, and noninvasive arterial blood pressure (NBP) modules was present in each patient's room (HP M1094B, Philips Medical Systems, N.A., Bothell, WA). The unit's central monitoring station was generally not staffed. Ventilators included a Siemens Servo 300/300A (Draeger Medical, Telford, PA), a Nellcor Puritan Bennett 840 (Nellcor Puritan Bennett LLC, Pleasanton, CA), or a Viasys Avea (VIASYS Healthcare, Conshohocken, PA). Alaris Medley infusion pumps were used in every room (Cardinal Health Dublin, OH). Flexiflo Quantum feeding pumps (Abbott Laboratories, Abbott Park, IL) were used in 13 observed rooms.

### 3.3.2 Data Recording

Time-stamped detailed information of alarms and the presence of health care team members were recorded manually using a COMPAQ iPAQ Pocket PC (Hewlett-Packard Company, Palo Alto, CA) and abcDB Database v.6.0 (PocketSOFT.ca, Lloydminster, SA, Canada). For health care team members, the time of entrance and exit as well as the provider category were recorded using a predefined list. When an alarm occurred, the observer recorded the device sounding the alarm, the alarm threshold settings, the alarm cause if identifiable, and the variable that produced the alarm: heart rate, Spo2, arterial blood pressure or NBP, pulmonary artery pressure, central venous pressure, temperature, peak airway pressure, minute volume (MV), tidal volume (TV), respiratory rate (RR) and apnea, InfP faults and feeding pump (FeedP) faults. For bedside tasks, the observer selected interventions from a predefined list and added free text comments with more detail. The following task categories were used: device alarm silenced, drug administered/dosage changed, patient assessment, physical therapy, washing, oral care, patient monitor settings changed, ventilator settings changed, data charted, arterial blood gas drawn, blood glucose levels measured, patient repositioned, airway suctioned, or other action taken.

### 3.3.3 Alarm Classifications

During the study, the observer classified each alarm as true, true irrelevant or false. However, the observer was not a clinician, so all alarms were reclassified after the

conclusion of the study using the following categories: effective, ineffective, or ignored. An alarm was classified as effective when an alarm-related action was performed by a qualified health care provider within 5 min of the end of the alarm. A qualified provider is one who has the authority to take alarm-related action. For example, physical therapists, phlebotomists, and health care assistants were only qualified to call for assistance, whereas nurses were qualified to administer medications, suction the patient's airway and change patient monitor settings. Only respiratory therapists and physicians were qualified to change ventilator settings.

Effective alarms were separated into two categories based on the action performed: (a) Technical actions include restarting infusion pumps, changing alarm thresholds, remeasuring values, changing sensor positions, reconnecting breathing circuits and all other equipment-related actions, and (b) patient actions included giving sedatives to an agitated patient, suctioning the airway, changing vasoactive drug infusion rates, repositioning agitated patients, and all other patient-related actions. An alarm was classified as ineffective if the alarm sounded, but a qualified health care provider did not enter the room in response to the alarm or was not present during the alarm. An alarm was classified as ignored when a qualified health care provider was present in the patient's room and no alarm-related action was taken during or within 5 min of the end of the alarm or the alarm was silenced from the nursing station and no action occurred.

### 3.3.4 Data Analysis

Analysis of the data was performed using MATLAB (The MathWorks, Natick, MA). The pocket PC generated ACCESS/EXCEL files (Microsoft Corporation, Redmond, WA) were parsed, events were categorized and alarm start and end times were paired with the times a person entered and left the room.

## 3.4 Results

Twenty-two health care team members participated in the study and gave written consent: 13 nurses, 3 nursing student interns, 3 respiratory therapists, 1 health care assistant, and 2 attending physicians. Several others, including phlebotomists, technicians and residents, who participated in the study gave verbal consent. Two-hundred

hr of data were collected from 22 patients over 24 days (13 males and 9 females, mean age 54.6 18.5 yr with a range from 21 to 93 yr). One day's data were lost and during 1 day participating health care team members did not care for a patient who met the inclusion criteria. Observations were made for an average of 9.16 hr per day (range, 6.25–10.5 hr). Two patients' lungs were ventilated using a Viasys Avea ventilator, 10 patients using a Siemens Servo 300 or 300A ventilator and 10 patients using a Nellcor Puritan Bennett 840 ventilator. Respiratory therapists, and occasionally the attending physicians or fellow physicians, changed the ventilator alarm thresholds; nurses changed the cardiac monitor alarm thresholds. We observed 10 changes to the patient monitor's alarm settings (5 NBP, 1 Spo2 and 4 not recorded) and 23 changes to ventilator alarm settings (8 MV, 4 peak airway pressure, 4 TV, 1 RR, 1 multiple changes, and 5 not recorded).

During the 200 hr of observation, 1214 alarms occurred (6.07 alarms per hr): Table 3.1 shows that 5.3% were effective and patient-related, 17.7% were effective and technically related, 36.3% were ineffective, and 40.7% were ignored. Figure 3.1 shows the number of alarms generated by each variable and the length of time each alarm was active. The median alarm length was 17 sec (range, 1 sec to 17.25 min): 45.1% lasted for 15 sec, 74.4% for 30 sec, and 89.4% for 60 sec. Of all the alarms, 34.3% ended without any health care team member being present in the patient's room. Thus they canceled themselves when the alarming condition cleared. Many more alarms cleared when no health care team member qualified to respond to this alarm was present. Only the feeding pump and the infusion pump always required user intervention for the alarm to stop. Figure 3.2 shows the total number of alarms for each of the four alarm types. A 19 sec alarm delay would reduce the number of ignored and ineffective alarms by 67.1%, whereas a 14 sec alarm delay would reduce it by 51.3%. For the effective alarms, the median time between the end of the alarm and the timestamp for the solution was 20 sec; 77 solutions were performed before the alarm had ended.

**Table 3.1**: Alarm frequency, duration, and classification

| | No. of alarms (#) | Alarm frequency (#/hr) | Alarm duration (sec/hr) | Effective patient (%) | Effective technical (%) | Ignored (%) | Ineffective (%) |
|---|---|---|---|---|---|---|---|
| Tidal volume | 247 | 1.24 | 15.9 | 7.7 | 3.6 | 39.3 | 49.4 |
| Minute volume | 197 | 0.99 | 21.0 | 9.1 | 7.1 | 55.8 | 27.9 |
| Pulse oximeter | 188 | 0.94 | 36.5 | 1.1 | 3.7 | 32.4 | 62.8 |
| Infusion pump | 147 | 0.74 | 42.7 | 0.0 | 82.9 | 17.1 | 0.0 |
| Heart rate and arrhythmias | 134 | 0.67 | 14.0 | 3.7 | 5.2 | 50.0 | 41.0 |
| Blood pressure (arterial and noninvasive) | 127 | 0.64 | 38.2 | 7.1 | 12.6 | 53.5 | 26.8 |
| Respiratory rate | 75 | 0.38 | 10.2 | 8.0 | 9.3 3 | 7.3 | 45.3 |
| Peak airway pressure | 37 | 0.19 | 2.9 | 13.5 | 2.7 | 43.2 | 40.5 |
| Other | 32 | 0.16 | 2.3 | 0.0 | 18.8 | 59.4 | 21.9 |
| Feeding pump | 30 | 0.15 | 13.7 | 0.0 | 90.3 | 9.7 | 0.0 |
| Overall | 1214 | 6.07 | 197.5 | 5.3 | 17.8 | 40.7 | 36.2 |

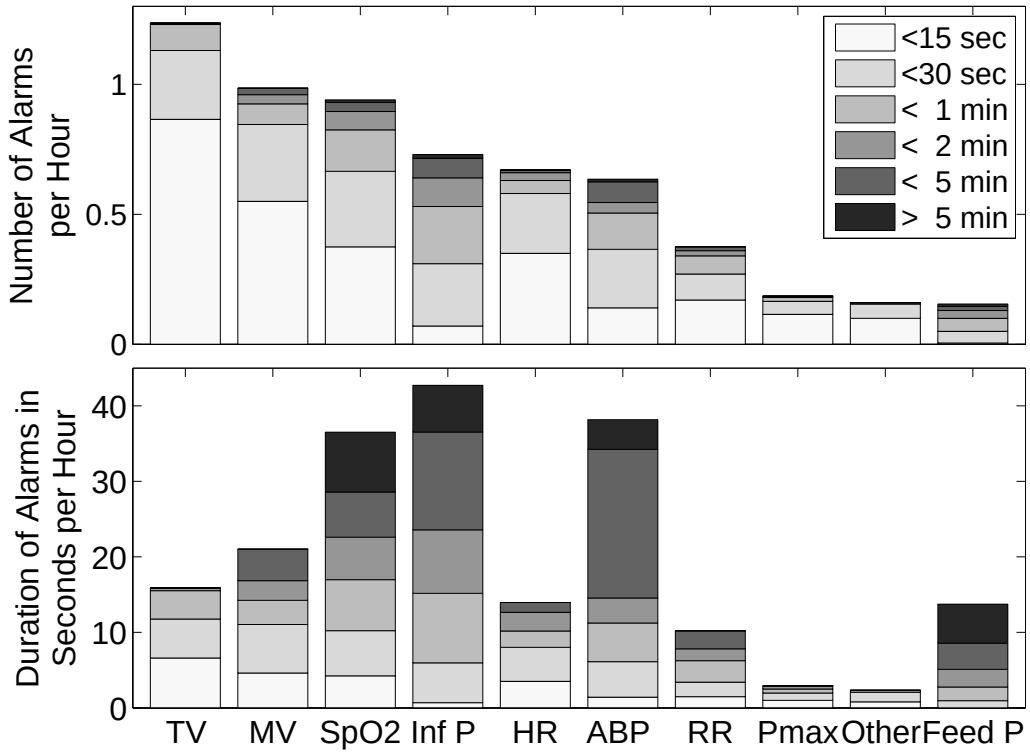


**Figure 3.1**: Number and duration of alarms per hr. The alarms are sorted by the alarm frequency, starting with the device with the most alarms per hr. The gray shading indicates the length the alarm was active, where each category does not include alarms already included in shorter-length categories. Alarms are: HR heart rate and arrhythmias; Spo2 pulse oximeter; ABP arterial or noninvasive blood pressure; Pmax peak airway pressure; MV minute volume; TV tidal volume; RR respiratory rate; InfP infusion pump; FeedP feeding pump; and Other all alarms not fitting into these categories.

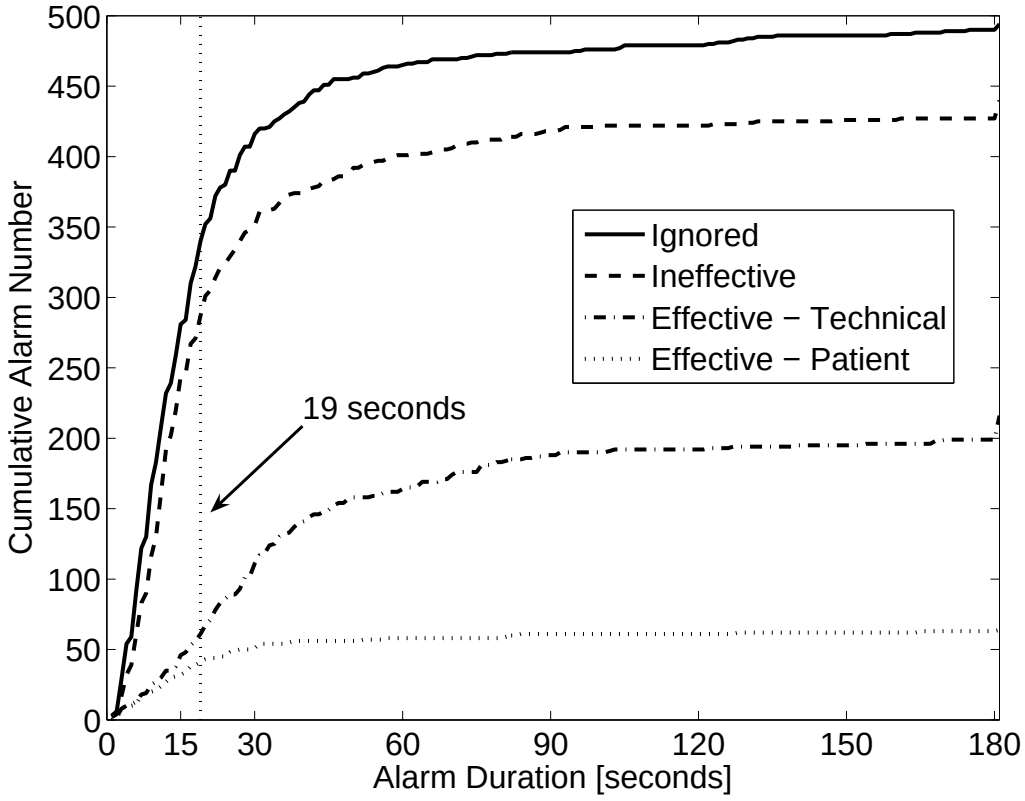


**Figure 3.2**: Cumulative alarm number and classification. Alarms lasting longer than 180 sec were categorized as having lasted 181 sec. The dashed line at the 19 sec mark indicates the proposed alarm delay duration.

### 3.4.1 Ventilator Alarms

Table 3.2 shows that ventilator manufacturers have taken different approaches for MV, TV and RR alarms. The Servo 300/300A ventilator does not alarm with TV or RR but with MV (TV times RR). The Nellcor 840 and Avea have separate alarms for all three related variables. As a consequence, the Nellcor 840 and Avea produced 4.29 alarms/hr and 5.43 alarms/hr, respectively, whereas the Servo 300 produced only 1.03 alarms/hr. However, while the percentages of ineffective and ignored alarms of the 3 ventilators (Servo 300, Nellcor 840 and Avea) were similar (83%, 84%, and 88%), our observation periods were not equal (94, 90, and 16 hr); therefore a statistical comparison was only performed between the Servo 300 and the Nellcor 840 group.

### 3.4.2 Unnecessary Alarms Occurring During Patient Care

During or within 2 min after suctioning, washing, repositioning, and oral care, 152 ineffective and ignored ventilator alarms were recorded (Table 3.3). A 2-min time window was chosen because the alarm silence button disabled alarms for 2 min. The primary alarm reason for 57 ventilator alarms, coded during the observation, was coughing. Patients' spontaneous breathing efforts were the cause for 118 ventilator alarms. Because no one was present in the room when 43.1% of the alarms started, they were not caused by a health care team member's actions.

### 3.4.3 Health Care Provider Presence and Tasks

During the 200-hr study period, 1271 separate entries by a health care team member to the room being observed were recorded; their average stay was 4.6 min (range, from 2 sec to 80.5 min). As seen in Figure 3.3, nurses made 65.7% of all visits; the patient's primary nurse made 44.8% of the visits. Of all providers, 15.6% stayed 30 sec and 70.8% 5 min. Nurses contributed to the longest duration of health care team members' stay in the patient's room (62.6%, primary nurse 37.7%). During the 200-hr study period 2344 tasks were performed (Fig. 3.4). On average, 11.7 tasks per hr were performed (range, 6.9 –21.5 task/hr), and most were done by the nursing staff. The most common tasks were nurses administering medications or changing infusion rates (2.3/hr), silencing alarms (1.3/hr), charting (1.1/hr), and patient assessments (0.7/hr).

**Table 3.2**: Number of ventilator alarms per hr

| | Tidal volume | Respiratory rate | Minute volume | Airway pressure | Apnea | Total number of alarms/hr |
|---|---|---|---|---|---|---|
| Servo 300 | | | 0.62 | 0.39 | 0.01 | 1.03 |
| Nellcor 840 | 2.39 | 0.59 | 1.23 | | 0.17 | 4.38§ |
| Avea | 2.04 | 1.40 | 1.72 | | 0.19 | 5.36 |

Average ventilator alarm thresholds were set to 36.6±6.6 mm Hg for peak airway pressure, 4.9±3.0 and 16.3±4.2 L/min for low and high minute volume, 37.6±7.1 L/min for respiratory rate, 283±78 and 954±144 mL for low and high tidal volume.

§ $P<0.005$ using a t-test with 2 tails and unpaired variance for alarms per day, normalized by duration of use during each day, between the Servo and the Nellcor group. No comparisons with the Avea group are reported because of its infrequent use during our study.

**Table 3.3**: Ventilator alarms occurring during or within 2 min of patient care tasks

| Task name | Minute volume alarms | Tidal volume alarms | Respiratory rate alarms | Peak airway pressure alarms | Apnea alarms | Inspired oxygen fraction (Fio2) alarms | Total number of alarms |
|---|---|---|---|---|---|---|---|
| Suctioning | 26 (21%) | 27 (13%) | 8 (19%) | 3 (8%) | 9 (47%) | 1 (100%) | 74 |
| Repositioning | 22 (18%) | 19 (9%) | 12 (28%) | 4 (11%) | 2 (11%) | 0 | 59 |
| Oral care | 6 (5%) | 2 (1%) | 1 (2%) | 0 | 0 | 0 | 9 |
| Washing | 4 (3%) | 1 (0%) | 2 (5%) | 3 (8%) | 0 | 0 | 10 |

The number in parenthesis is the percentage of the total alarms in each alarm category

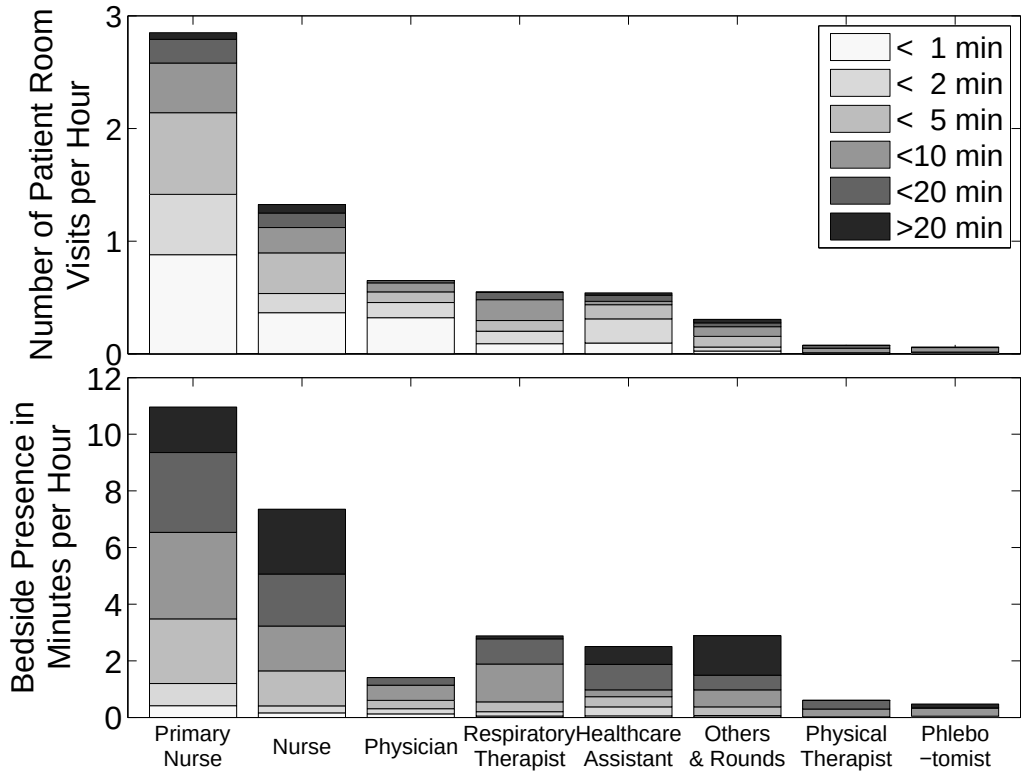


**Figure 3.3**: Number and duration of health care provider visits to the patient's room. The providers are sorted by the number of visits, starting with the most visits per hr. Physicians include attending physicians, fellow physicians, and resident physicians. The gray shading indicates the duration a health care provider stayed in the patient's room. Each duration category does not include durations already included in shorter-length categories.

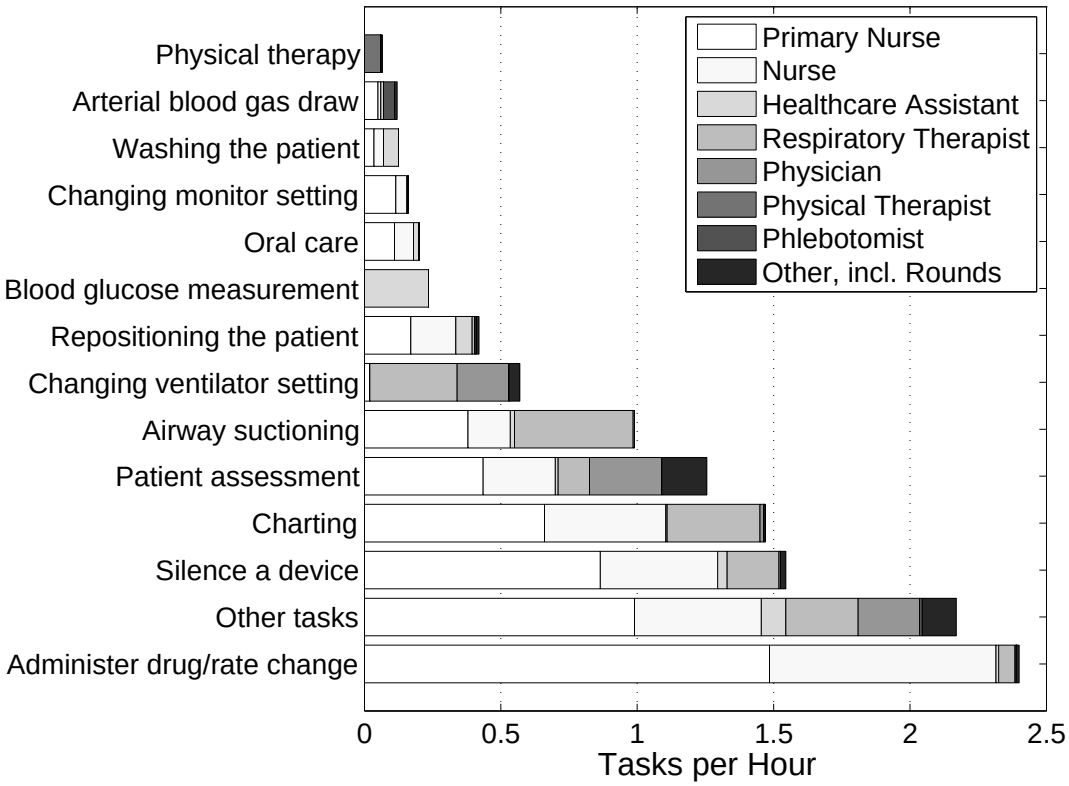


**Figure 3.4**: Tasks frequency. The number of tasks completed each hr, sorted by frequency. Physicians include attending physicians, fellow physicians, and resident physicians.

## 3.5 Discussion

Of the 1214 alarms that occurred during our 200-hr observation period, only 23% were effective. If the alarm onset would be delayed for 19 sec, two-thirds of the ignored and ineffective alarms could have been avoided (Fig. 3.2). Suctioning, washing, repositioning, and oral care caused 152 ineffective and ignored ventilator alarms, of which 33 were longer than 19 sec. If the alarm systems had been contextually aware of patient care procedures and waited 19 sec before sounding an alarm, the combined ineffective and ignored alarm rate could have been reduced from 934 (77%) to 274 (50%) and the total number of alarms reduced by 54%. This reduction in the number of alarms would be clinically relevant as alarm noise has a detrimental effect on patients' sleep and ICU outcome.[12] Additionally, this reduction should reduce alarm fatigue, a problem commonly observed in ICUs.[4, 13]

### 3.5.1 Comparison with the Literature

Our observation of 6.1 alarms/hr is consistent with a literature review by Imhoff and Kuhls[1] reporting 1.6 to 14.6 alarms/hr. We did not classify alarms as false and true; however, 77% of our alarms were ineffective and ignored alarms, which is similar to the false alarm rate of 90% reported by Imhoff and Kuhls.[1]

### 3.5.2 Alarm Classification Method

Tsien and Fackler[3] define true, true irrelevant, and false alarms as: "True Positive, Clinically Relevant was used to indicate the monitoring device sounded an alarm, the alarm was appropriate given the actual data value as compared with the set threshold value, and the patient's condition required prompt attention. ... True Positive, Clinically Irrelevant was used to indicate the monitor sounded an alarm, the alarm was appropriate given the input data value as compared with the set threshold value, but the patient's condition had not changed in a way that required additional medical attention. ... False Positive was used to indicate that the monitor sounded an alarm, but the alarm was inappropriate given the input data value. ... The alarm was false because the reported value did not reflect the patient condition."

Classifying alarms into effective, ineffective, and ignored alarms has three advantages over the traditional method using true, true irrelevant, and false alarms: (a)

the alarm classification can be performed by a trained observer rather than an expert clinician; (b) the classification can be performed after the completion of the study, as long as tasks and providers' actions are recorded; and (c) the criteria using a time cutoff and a task requirement related to the alarm makes it more objective than a single clinician's decision. However, there are also three disadvantages associated with this approach: (a) our method departs from the current alarm study literature using true and false alarms,[1] (b) an effective alarm might be misclassified as an ineffective or ignored alarm if the response takes longer than 5 min to initiate, and (c) the alarm records must include tasks and health care provider actions.

### 3.5.3 Introducing an Alarm Delay

A delay would improve alarm reliability at the expense of lengthening the response time. It seemed that the staff currently respond selectively to alarms or wait before responding. They went to the patient's room in response to only 9.1% of the alarms, yet of these alarms 69.4% were effective alarms. It would be better for the alarm system to automatically introduce a delay rather than relying on the busy clinician to keep track of alarm duration. This proposal is consistent with a pulse oximeter study in which a 6 sec delay reduced the alarm rate by 50%.[12] Waiting 19 sec before sounding an alarm would have reduced the number of Spo2 alarms by 52%. Newer Spo2 monitors claim to have reduced the false alarm rate to 15% with only minor delays by using better signal processing techniques.[14, 15] However, to keep the patient safe, asystole and ventilator disconnect/apnea alarms should be exempt from this delay.

### 3.5.4 Reducing Ventilator Alarms

TV was the most frequently occurring alarm; MV was the second most frequent (Table 3.1). The TV signal from the ventilator is a noisy signal, especially in patients with spontaneous breathing efforts and with active airway protection reflexes (coughing). TV alarms frequently occurred after suctioning. Waiting 19 sec to announce a low TV would have had very little consequence to the patients we observed, as they all had MV and blood oxygenation saturation alarms that sound before desaturation occurs. The 19 sec delay would have reduced the number of TV

alarms in our observation period by 18% and the number of MV alarms by 37%. Apart from the patients' spontaneous respiratory efforts and coughing after suctioning, the leading causes for ventilator alarms were the lack of adaptations of the alarm threshold when ventilation modes were changed.

Perhaps TV and RR alarms are not necessary and an MV alarm is sufficient. The Servo 300, with an MV alarm, produced only 1.03 alarms/hr, whereas the Nellcor 840 and Avea, with RR, TV and MV alarms, produced 4.29 alarms/hr and 5.43 alarms/hr, respectively (Table 3.2). However, without tracking patient outcome, we cannot say which strategy is best.

An even more conservative approach is taken by Philips with their Intellivue Event Surveillance system in which both MV and Spo2 must cross alarm thresholds before an event is identified (Philips Medical Systems, Andover, MA). This approach has been well accepted in neonatal care units. Our proposed approach would significantly improve the reliability of ventilator alarms and may result in more timely attention when the patient is truly at risk.

### 3.5.5 Reducing InfP and FeedP Alarms

InfP alarms were the longest in duration. One possibility to explain this behavior is that most InfP alarms did not identify a critical event, when the patient was in danger, and therefore the staff tended to ignore them for longer periods of time. InfP alarms had a high effective alarm rate (83%) because an alarm, once triggered, does not stop until the technical problem is resolved. We observed the nursing staff intentionally entering a smaller infusate volume than was available, so that the InfP alarm reminded them when the pump was nearly empty. Such alarm tailoring indicates the need for alarm redesign.[16] Here manufacturers could implement a lower priority reminder function to support this behavior. In general, InfP alarms signaled mechanical failures and empty infusates, rather than patient trouble, and should be used only in situations involving the delivery of a life-supporting drug.

The FeedP alarms had a high effective alarm rate (90.3%) because there was no alarm silence button and the technical problem had to be fixed before the alarm would stop.

### 3.5.6 Reducing Alarms Occurring During Patient Care

Nursing care seems to generate a significant number of alarms. During our observations, 57% of the alarms occurred when a health care team member was in the room. Considering that nurses were in the room for only 18.3 min an hr, a disproportionate number of alarms occurred while they were in the room.

If the 2-min alarm silence button had been activated before suctioning, 74 alarms would have been prevented and the ventilator alarm rate would decrease from 2.8 alarms/hr to 2.5 alarms/hr.

Washing caused 10 unnecessary alarms; repositioning caused 59. If repositioning the patient were automatically detected by a mattress detection system,[17, 18] the number of unnecessary ventilator alarms could have been reduced by approximately 10%.

### 3.5.7 Health Care Provider Presence and Tasks

Figure 3.4 shows that silencing alarms constitutes approximately 16% of a nurse's bedside tasks. A radio frequency identification tracking system[19] that could identify when a nurse arrives in the patient room and automatically silence alarms could reduce workload. However, caution is needed when changing the way alarms function because the new function might lead to unintended consequences known as "automation surprises."[20] Noise pollution could be reduced if alarms were to sound or be visually signaled outside the patient's room when a provider was not present in the room.

## 3.6 Conclusions

The number of ignored and ineffective alarms in a MICU could decrease from 934 to 274 by introducing a 19 sec alarm delay, and by automatically detecting suctioning, patient repositioning, oral care, and blood gas sampling. Hopefully, more reliable alarms will elicit a more timely response, reduce workload, reduce noise pollution, and potentially improve patient safety.

## 3.7 Acknowledgments

The authors would like to thank all participating health care team members at the University of Utah Hospital for their support.

## 3.8 References

1. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg. 2006 May;102(5):1525–1537.

2. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med. 1999 Dec;25(12):1360–1366.

3. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med. 1997 Apr;25(4):614–619.

4. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med. 1994 Jun;22(6):981–985.

5. Quinn ML. Semipractical alarms: a parable. J Clin Monit. 1989 Jul;5(3):196–200.

6. Rheineck-Leyssius AT, Kalkman CJ. Influence of pulse oximeter settings on the frequency of alarms and detection of hypoxemia: Theoretical effects of artifact rejection, alarm delay, averaging, median filtering or a lower setting of the alarm limit. J Clin Monit Comput. 1998 Apr;14(3):151–156.

7. Bricon-Souf N, Newman CR. Context awareness in health care: A review. Int J Med Inform. 2006 Feb;.

8. Dey AK. Providing Architectural Support for Building Context-Aware Applications [PhD thesis]. Georgia Tech. Atlanta, GA; 2000.

9. O'Carroll TM. Survey of alarms in an intensive therapy unit. Anaesthesia. 1986 Jul;41(7):742–744.

10. Koski EM, Mäkivirta A, Sukuvaara T, Kari A. Frequency and reliability of alarms in the monitoring of cardiac postoperative patients. Int J Clin Monit Comput. 1990 Apr;7(2):129–133.

11. Seagull FJ, Sanderson PM. Anesthesia alarms in context: an observational study. Hum Factors. 2001 Spring;43(1):66–78.

12. Christensen M. Noise levels in a general intensive care unit: a descriptive study. Nurs Crit Care. 2007 Jul;12(4):188–197.

13. Meredith C, Edworthy J. Are there too many alarms in the intensive care unit? An overview of the problems. J Adv Nurs. 1995 Jan;21(1):15–20.

14. Petterson MT, Begnoche VL, Graybeal JM. The effect of motion on pulse oximetry and its clinical significance. Anesth Analg. 2007 Dec;105(6 Suppl):S78–84.

15. Kawagishi T, Kanaya N, Nakayama M, Kurosawa S, Namiki A. A comparison of the failure times of pulse oximeters during blood pressure cuff-induced hypoperfusion in volunteers. Anesth Analg. 2004 Sep;99(3):793–796.

16. Watson M, Sanderson P, WJ R. Tailoring reveals information requirements: the case of anaesthesia alarms. Interacting with Computers. 2004 Apr;16(2):271–293.

17. Miyamoto Y, Yonezawa Y, Maki H, Ogawa H, Hahn AW, Caldwell WM. A system for monitoring heart pulse, respiration and posture in bed. Biomed Sci Instrum. 2002;38:135–138.

18. Watanabe K, Watanabe T, Watanabe H, Ando H, Ishikawa T, Kobayashi K. Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method. IEEE Trans Biomed Eng. 2005 Dec;52(12):2100–2107.

19. Goldman JM, Stahl EJ, Sandberg WS, Fairbrother PL, Atkins PL, Egan MT, et al. Automatic Identification of the "Start of Anesthesia Care" using Indoor Positioning System (IPS) Derived Spatial-Temporal Association. In: Proceedings 2004 STA Annual Meeting; 2004. .

20. Cook RI, Woods DD. Adapting to new technology in the operating room. Hum Factors. 1996 Dec;38(4):593–613.

# CHAPTER 4

# A TOOL PREDICTING MEAN ARTERIAL BLOOD PRESSURE VALUES IMPROVES THE TITRATION OF VASOACTIVE DRUGS*

## 4.1 Abstract

Vasoactive drug infusion rates are titrated to achieve a desired effect, e.g., mean arterial blood pressure (MAP), rather than using infusion rates based on body weight. The purpose of this study is to evaluate a method to automatically identify a patient's sensitivity to sodium-nitroprusside, dobutamine or dopamine and to evaluate, whether an advisory system that predicts MAP 5 min in the future enhances a clinician's ability to titrate sodium-nitroprusside infusions.

We used published models implemented in MATLAB to simulate the response of 100 individual patients to infusions of sodium-nitroprusside, dopamine and dobutamine. The simulated patient's sensitivity to the three drugs was identified using an adaptive filter approach, where MAP was altered in a binary stepwise fashion. Next, 9 nurses were asked to control the MAP of 6 of the simulated patients. For half of the patients, we used the identified sensitivity to predict and display MAP 5 min into the future.

Identifying each individual patient's sensitivity improved the accuracy of the MAP prediction by 75% for sodium-nitroprusside, 82% for dopamine and 52% for dobutamine over the MAP prediction based on an "average" patient's sensitivity.

*With kind permission from Springer Science+Business Media: Görges M, Westenskow DR, Kück K, Orr JA. A tool predicting future mean arterial blood pressure values improves the titration of vasoactive drugs. J Clin Monit Comput. 2010. DOI: 10.1007/s10877-010-9238-0. 

The advisory system shortened the median time to reach the desired MAP from 10.2 to 4.1 min, decreased the median number of infusion rate changes from 6 to 4, and resulted in a significant reduction of mental workload and effort.

Patient-specific drug sensitivity identification significantly improved the prediction of future MAP. By predicting and displaying the expected MAP 5 min in the future, the advisory system helped nurses titrate faster, reduced their perceived workload and might improve patient safety.

## 4.2 Introduction

Vasoactive drug infusion rates[1–4] are titrated to achieve a desired effect,[3] rather than using standardized doses, because the interpatient variability in response to the drug varies significantly. Current practice is to start with an initial infusion rate of approximately 10-25% of a typical infusion rate for a drug and observe the change in mean arterial blood pressure (MAP). Once the MAP looks stable, the dose is increased or decreased until the MAP goal is reached. Because there is a time delay between the change in an infusion rate and the subsequent change in the patient's MAP, considerable nursing or physician time is often required to titrate the infusion rate until the desired effect is achieved.

### 4.2.1 Alternatives to Manual Titration

Syroid et al.[5] and Drews et al.[6] used pharmacologic models to predict drug effects (sedation, analgesia and muscle relaxation) 10 min in the future. When their advisory system was used, physicians better controlled their targeted drug effects, especially towards the end of the procedure, which resulted in patients waking up more than 2 min earlier.[6] A nonmedical application is the flight glide path adviser,[7] which enhances a pilot's performance by predicting where the plane will be in relation to the runway a few minutes in the future. DeLucia et al.[8] state that predictive displays of patient information that help nurses anticipate the short-term future states of patients would be a particularly useful technology. Such a system could decrease nurses' and physicians' workload by allowing faster titration and could potentially improve patient safety by freeing up nurses' time for other tasks.

### 4.2.2 Purpose of the Study

The purpose of this study is to use simulation to test the feasibility of using small step changes in infusion rate to automatically and continuously identify a patient's sensitivity to sodium nitroprusside (SNP), dobutamine or dopamine, as the drug is being infused. The second purpose is to evaluate whether an advisory system that predicts MAP 5 min in the future enhances a clinician's ability to manage SNP infusions.

## 4.3 Methods

To predict MAP 5 min into the future (as seen in Figure 4.1a) the change in MAP caused by the SNP infusion ($I_{SNP}(t)$ ) needs to be calculated. One simple way, which includes an infusion delay and models the patient's delayed response to SNP like a low pass filter, is

$$MAP\,(t) = MAP_{Baseline} + K_{SNP} \cdot ((1-\alpha) \cdot I_{SNP,filtered}\,(t-1) + \alpha \cdot I_{SNP}\,(t - t_{Delay})) \tag{4.1}$$

where $MAP_{Baseline}$ is the MAP without SNP infusion, $K_{SNP}$ is the patients sensitivity to SNP, $\alpha$ is the strength of the low-pass filter and $t_{Delay}$ is the infusion delay. However, due to the large variability in patient's sensitivity to SNP, it is necessary to identify the individual's sensitivity (see Figure 4.2) in order to provide a precise prediction.

### 4.3.1 Identification of Patient Sensitivity

We identify the patient's sensitivity as the drug was being delivered by increasing and decreasing the infusion rate in a binary stepwise fashion and measuring the corresponding changes in MAP. The following algorithm, with the goal of minimizing the fit between the recorded infusion rate and the measured MAP, was used to identify the patient's sensitivity to SNP (see Figure 4.2):

1. Begin with an initial infusion transport delay ($T_i$) of 45 sec and choose a random initial value for the filter strength $\alpha$ between 0.003 (corresponding to a of 370 sec) and 1 (no filter).


**Figure 4.1**: Sodium-nitroprusside titration advisor. There are five main elements: 1) A 30 min trend of mean arterial blood pressure (MAP) and drug infusion rate, 2) Numeric values for the current MAP and infusion rate, 3) A slider next to the infusion rate trend to change the infusion rate and 4) A numeric keypad to set the infusion rate and 5) the patient's sensitivity in the upper right corner. The horizontal dashed line indicates the current time. The plot to the right of this line shows predicted values for MAP 5 min in the future. The top panel shows the advisory system, when the prediction of future MAP is available. The bottom panel shows the advisor, when predicted pressure and estimated sensitivity are not available.

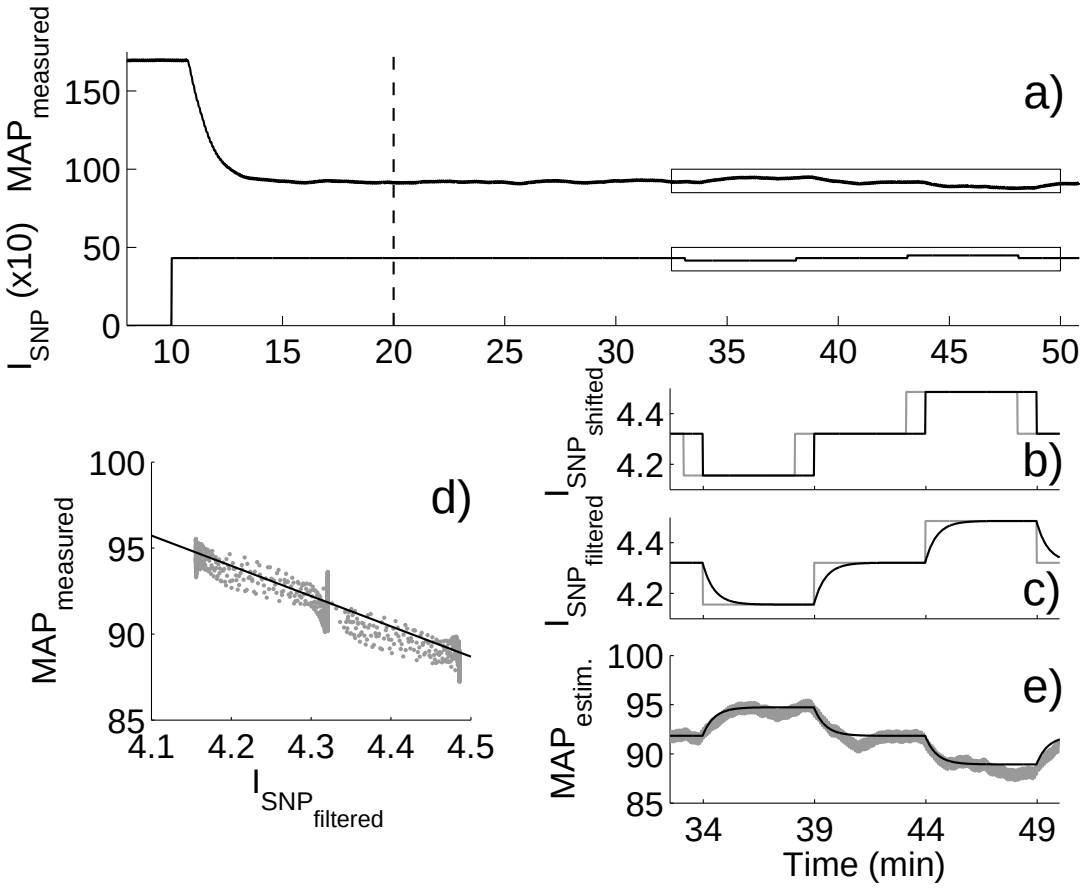


**Figure 4.2**: An illustration of the steps in the sensitivity identification algorithm: a) mean arterial blood pressure (MAP) in mmHg and sodium-nitroprusside (SNP) infusion rate in $\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$. The identification is performed in the area indicated by the boxes. b) The gray line shows an expanded view of the SNP infusion rate in panel a. The black line shows the SNP infusion rate, shifted to the right by the infusion delay. c) The gray line shows the shifted SNP infusion rate from panel b. The black line shows the shifted and low-pass filtered infusion rate. d) Shows the linear fit between the filtered infusion rate from panel d and the measured MAP from the boxed area in panel a. e) The gray line shows the measured MAP from the boxed area in panel a. The black line shows the predicted MAP calculated using the patient's sensitivity to SNP, as derived from the slope in panel d, and the baseline MAP, as derived from the y-intercept in panel d.

2. Shift the original infusion rate $I_{Drug}(t)$ (Figure 4.2a) forward in time by $T_i$ using $I_{shifted} = I_{Drug}(t - T_i)$ so that it is more closely aligned in time with the corresponding changes in $MAP_{measured}(t)$ (Figure 4.2c).

3. Apply a low-pass infinite impulse response (IIR) filter to calculate a filtered drug infusion rate ($I_{filtered}(t)$ ): $I_{filtered}(t) = (1 - \alpha) \cdot I_{filtered}(t - 1) + \alpha \cdot I_{shifted}(t)$ to make its shape the same as the MAP curve, where $\alpha$ is a value between 0 and 1.0 (Figure 4.2d).

4. Use a linear fit between $I_{filtered}(t)$ and $MAP_{measured}(t)$ to calculate the MAP without drug $MAP_{baseline}$ (from the "y axis" intercept) and drug sensitivity ($K_{Drug}$) (from the slope). (Figure 4.2b).

5. Use $I_{filtered}(t)$, $K_{Drug}$ and $MAP_{baseline}$ to predict values for MAP: $MAP_{predicted}(t) = MAP_{baseline} + K_{Drug} \cdot I_{filtered}(t)$, and calculate the root mean square error (RMSE) between $MAP_{predicted}(t)$ and $MAP_{measured}(t)$ (Figure 4.2e).

6. Select a new $\alpha$ and repeat steps 1-4 until MATLAB's fminbnd optimization (golden section search) reports that $RMSE(t) - RMSE(t - 1) < 10^{-4}$ or the number of iterations exceeds 500.

7. Increase and decrease $T_i$ by 1 sec, repeat steps 1-5 for each $T_i$ to find the direction of $T_i$ resulting in the lower RMSE. Continue in the direction that lowers RMSE, limited to $T_i$= 0-80 sec, and repeat steps 1-5 until RMSE increases.

8. Finally, the patients sensitivity ($K_{Drug}$), the infusion delay ($T_i$) and filter strength ($\alpha$) that result in the lowest RMSE are reported. The obtained $K_{Drug}$, $T_i$ and $\alpha$ can now be used to predict $MAP(t)$ 5 min into the future, e.g., using Equation 4.1.

### 4.3.2 Sensitivity Identification Performance Evaluation

To measure the accuracy of the sensitivity estimation algorithm we simulated the blood pressure response to a SNP in 100 individual simulated patients and calculated the estimation error (the difference between the sensitivity identified by the estimation

algorithm and the sensitivity parameter used to simulate the individual patient's response).

#### 4.3.2.1 Creating Unique Patient Responses to SNP Infusions

Figure 4.3 shows the computer model, proposed by Slate[9] that we used to predict the change in a patient's MAP ($\Delta MAP(s)$ ) when SNP was given by intravenous infusion ($I_{SNP}$) :

$$\Delta MAP(s) = \frac{K_{SNP} \cdot e^{-T_i \cdot s} \cdot \left(1 + \beta \cdot e^{-T_{cr} \cdot s}\right)}{1 + \tau \cdot s} \cdot I_{SNP}(s) \tag{4.2}$$

where $s$ is the independent variable (Laplace transformed time). The model was implemented in MATLAB (The MathWorks Inc., Natick, MA). The model's parameters are listed in Table 4.1.[9]

The blood pressure transducer was modeled using by a low-pass filter to calculate the measured MAP: $MAP_{measured}(s) = 1/(4s+1) \cdot MAP(s)$. Slate[9] added two sources of noise: a) Second order 30 Hz low-pass filtered white noise with an amplitude ranging from 5-10 mmHg and b) A sinusoidal fluctuation of 2-4 mmHg to simulate the change in MAP with respiration (rate 6-12/min, random initial phase of 0-10 sec). A renin-angiotensin reflex was added when the MAP fell below 63.3 mmHg: $\Delta MAP_{renin}(s) = \frac{K_r \cdot e^{-T_r \cdot s}}{1 + s \cdot \tau_r} \cdot MAP(s)$, where $T_r$ is the activation time delay (60 sec), $K_r$ is the reflex gain (2 mmHg), $\tau_r$ is the reflex time constant (4 min) and $MAP(s)$ is the measured MAP. The reflex was limited so that it could not increase MAP by more than 35 mmHg.

Each of the simulated patients had a unique sensitivity to the drug and unique responses to changes in its infusion rate. This was performed by randomly selecting values for each of the model parameters listed in Table 4.1 . The value for the $i^{th}$ parameter and the $k^{th}$ patient was:

$$P_i(k) = Mean_i + SD_i \cdot Randn_{i,k} \tag{4.3}$$

where $k$=1-100, $i$=1-7 and $Randn$ is a number taken randomly from a normal distribution where the distribution had a mean of 0 and a standard deviation of 1. Times were rounded to the nearest sec.

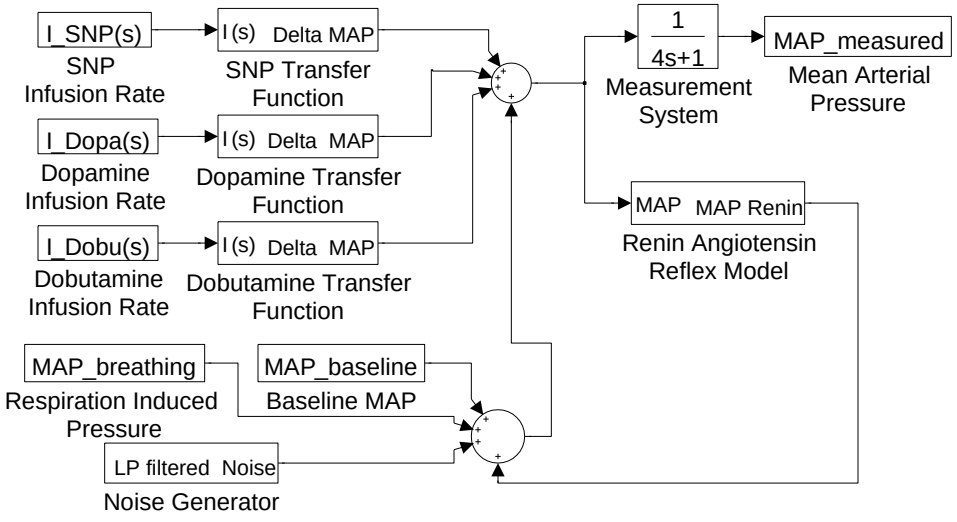


**Figure 4.3**: MATLAB implementation of Slate's sodium-nitroprusside model that includes two sources of noise, the baseline mean arterial pressure, the blood pressure measurement transducer and the renin-angiotensin reflex. We added dopamine and dobutamine transfer functions, converting an infusion rate into a mean arterial pressure (MAP) change, to the upper summation node. The three infusions were simulated one at a time.

**Table 4.1**: Sodium-nitroprusside model parameters

| Parameter | Description | Symbol | Mean±Standard Deviation | Limiting Range |
|---|---|---|---|---|
| $P_1$ | Patient's sensitivity | $K_{SNP}$ | $20 \pm 10 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ | $6.5 - 171 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ |
| $P_2$ | Recirculation coefficient | $\beta$ | $0 \pm 0.2$ | $0 - 40\%$ |
| $P_3$ | Infusion delay | $T_i$ | $40 \pm 10\,\text{sec}$ | $20 - 60\,\text{sec}$ |
| $P_4$ | Recirculation delay | $T_c$ | $45 \pm 7.5\,\text{sec}$ | $30 - 75\,\text{sec}$ |
| $P_5$ | System time constant | $\tau$ | $40 \pm 5\,\text{sec}$ | $30 - 60\,\text{sec}$ |
| $P_6$ | Baseline MAP | $MAP_{baseline}$ | $168 \pm 19\,\text{mmHg}$ | - |
| $P_7$ | Target MAP | $MAP_{goal}$ | $90 \pm 12.5\,\text{mmHg}$ | $> 70\,\text{mmHg}$ |

Parameters used in the Sodium-Nitroprusside transfer function (Equation 4.1). Values for the parameters $P_1 - P_5$ were obtained from Slate's PhD Thesis,[9] $P_6$ from Devlin et al.,[10] and $P_7$ was picked to be in the range of "textbook" normal blood pressures.

#### 4.3.2.2 Identification of SNP Sensitivity

For each of the 100 simulated patients, we used the patient's $K_{SNP}(k)$ and the recirculation factor ($\beta(k)$) to calculate the SNP infusion rate needed to lower the MAP from the baseline ($MAP_{baseline}(k)$) to the target ($MAP_{goal}(k)$). The baseline MAP was selected using Equation 4.3 with a mean and SD of 168±19 mmHg, which is a typical value observed in patients requiring SNP infusion, before the administration of the drug.[10] The target MAP was selected using Equation 4.3 with a mean and SD of 90±12 mmHg. After infusing SNP at this rate and waiting for 10 min to obtain steady state, we made three step-increase/step-decrease changes in the infusion rate to change MAP by ±3 mmHg. Each step was held for 5 min. The optimal number of steps, step size, and step duration were identified using simulations (see Appendix A).

The estimation algorithm used the response data from these three steps to identify the patient's sensitivity. We compared the sensitivity identified by the estimation algorithm $K_{SNP,estimated}(k)$ with the patient's sensitivity used to generate the response $K_{SNP}(k)$. Finally, we compared the predicted MAP change caused by a $2\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$ SNP infusion rate when given to each of the 100 simulated patients with the MAP change when given to a patient with a typical sensitivity of $20\,\frac{\mathrm{mmHg}}{\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}}$.

### 4.3.3 Dopamine and Dobutamine Sensitivity Identifications

We applied our sensitivity identification method to Dopamine and Dobutamine, both of which are used very frequently to manage MAP. For Dopamine we used a well-established model from the literature.[11] For Dobutamine, where such a model does not exist, we propose a new model based on observations in canines.[12] In both cases we evaluated the performance of the proposed sensitivity estimation algorithms in a similar fashion as for SNP (see Appendices B and C).

### 4.3.4 Blood Pressure Titration Tool

Approval from the University of Utah Health Sciences Center's institutional review board (ClinicalTrials.gov identifier: NCT00714012) was obtained. Nine medical intensive care unit nurses participated in the study. We used a 2 x 3 repeated-measures

within-subject experimental design, with the availability of the advisory system as the independent variable and the scenario order as the controlled variable.

#### 4.3.4.1 Apparatus

The titration screen shown in Figure 4.1 displayed a 30 min trend of MAP and SNP infusion rate, numeric values for the current MAP and infusion rate, and a slider and numeric keypad to change drug infusion rate. The target MAP was shown by a green zone and given as a numeric value in the instruction section of the screen. The advisory system, available in 3 of the 6 simulations, displayed the estimate of the patient's sensitivity to SNP and a prediction of the MAP 5 min into the future (Figure 4.1a). If the participant selected a MAP level, which could be different from the target MAP, it calculated the infusion rate required to reach this MAP level. To identify the patient's sensitivity, a SNP infusion rate of $1\,\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ was administered and the patient's sensitivity to SNP identified using three $0.1 \pm \text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ steps 5 min in duration.

#### 4.3.4.2 Training

In a training session, participants were shown the advisor in Figure 4.1, asked to increase the infusion rate, observe the delay between the increase in the infusion rate and the decrease in MAP, and use the advisor to titrate MAP to a target pressure.

#### 4.3.4.3 Scenario

Participants were asked to titrate a SNP infusion rate to reach a target MAP in 6 simulated patients. The 6 patients were randomly selected from a set of 200 patients generated using the methods previously described. The availability of the advisor in 3 of the 6 simulated patients was randomized for each of participant. Target pressures were selected randomly using Equation 4.3 with a mean and SD of 105±12.5 mmHg. Each simulated patient and target MAP was used only once. To increase workload the simulations were run at 3.6 times real time. Participants were encouraged to reach the target MAP in the shortest amount of time using the least number of infusion rate changes. They were asked to spend 100% of their attention on the task.

#### 4.3.4.4 Procedure

We recorded the time to reach each titration goal and the number of infusion rate changes. We also counted the number of times the infusion rate was changed when the MAP was below 70 mmHg or when the MAP was more than 10 mmHg below the target MAP. A "finished I am done" button advanced the simulation to the next patient if: a) the measured MAP was within 4mmHg of the target, b) the infusion rate was within $0.15\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$ of the required infusion rate needed to reach the target MAP and c) the infusion rate was not changed for 8.3 sec (30 sec of simulated time). These requirements prevented the participants from "cheating" by pressing the finished button as the MAP passed transiently through the target range. After each scenario, the participant completed a NASA Task Load Index (TLX) questionnaire.[13] The experiment was performed in the break room of the ICU, providing realistic ambient noise.

### 4.3.5 Data Analysis

Data were analyzed using MATLAB, where a repeated measure Friedman's ANOVA ($\alpha = 0.05$) was used to analyze the titration times, the number of infusion rate changes and NASA-TLX scores and Fisher's exact test was used to analyze both numbers counting episodes of low MAPs.

## 4.4 Results

### 4.4.1 Sensitivity Identification Performance

Figure 4.4 shows that the patients' sensitivity to SNP must first be identified before future values for MAP can be accurately predicted while SNP is being infused. When 100 simulated patients were given SNP at a typical infusion rate ($2\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$) the prediction of MAP 5 min in the future had an error of $-9.0 \pm 18.5\,\mathrm{mmHg}$ (mean±SD) if we assumed that all patients had the same population based sensitivity to SNP ($20\,\frac{\mathrm{mmHg}}{\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}}$). When each individual patient's sensitivity was estimated using our algorithm, the error was $-1.8 \pm 4.1\,\mathrm{mmHg}$. The MAP prediction was 75% more accurate when the individual patient's sensitivity was identified. Figure 4.5 shows the difference between the simulated patients' actual sensitivity to SNP and the sensitivity estimated using our algorithm. After making three step changes in the SNP

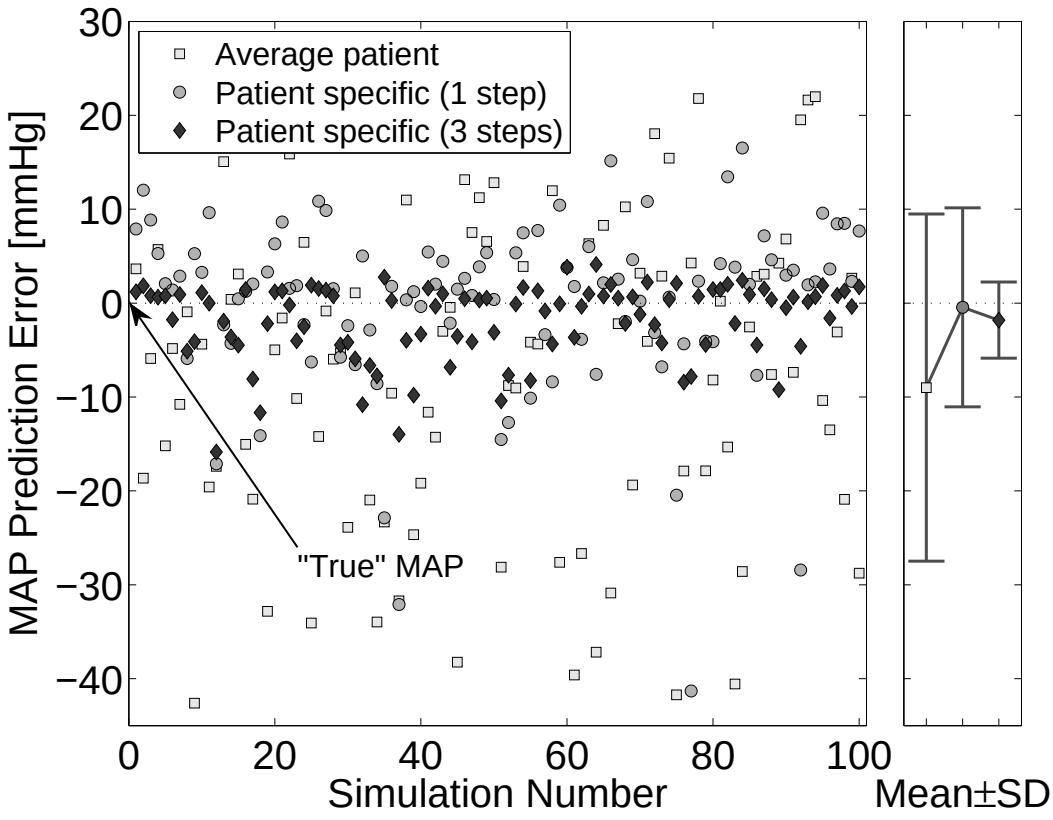


**Figure 4.4**: The error in our prediction of mean arterial blood pressure (MAP) 5 min after starting a sodium-nitroprusside (SNP) infusion rate of $2\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$ (at steady state): The open boxes show the error when the prediction algorithm uses an "average" patient's sensitivity of $20\,\mathrm{mmHg}/\left(\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}\right)$. The gray circles show the error using the patient's sensitivity identified after making one infusion rate step change. The black diamonds show the error after three infusion rate step changes. The bars in the right panel show mean±standard deviation of the prediction error.

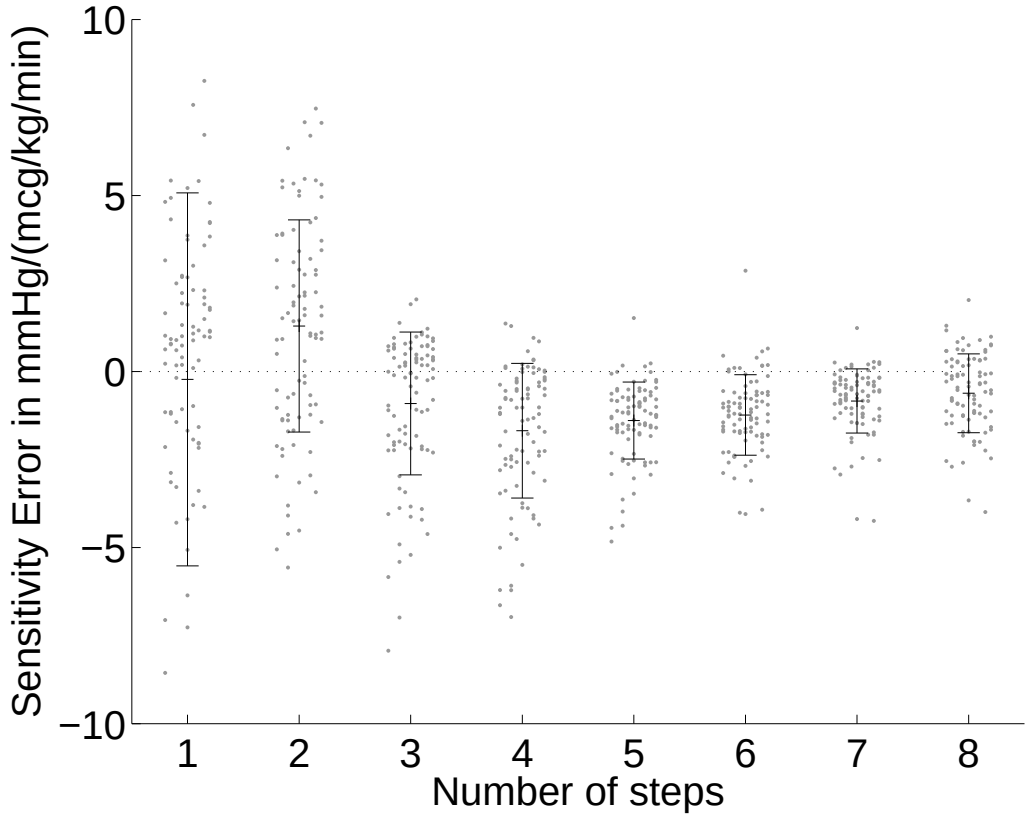


**Figure 4.5**: The error in our estimation of sensitivity to sodium-nitroprusside (SNP) for 100 simulated patients. The error decreased with each of the first five step changes in the SNP infusion rate, with each step lasting 5 min. The bars show the mean±standard deviation of the error at each step.

infusion rate, where each step resulted in a 3 mmHg change in MAP lasting 5 min, the difference between the estimated and actual sensitivity had a mean value and standard deviation of $-0.91 \pm 2.03 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$. The error in the estimation of infusion delay was 3±9 sec. There is a trade-off between identification time and estimation error: a 5 min identification time results with an error of $-0.22 \pm 5.30 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ whereas a 25 min identification time results in an error of $-1.39 \pm 1.10 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$.

Appendices B and C show the results for dopamine and dobutamine: For dopamine (see Figure 4.6) the mean the difference between the estimated and actual sensitivity, after three 15 min steps in the infusion rate, was $1.12 \pm 0.85 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$. For dobutamine (see Figure 4.7) the mean difference after a single 10 min step was $-1.88 \pm 1.99\,\text{mmHg}$. Sensitivity estimation resulted in an 82% improvement in the MAP prediction for Dopamine and a 52% improvement in the MAP prediction for Dobutamine.

### 4.4.2 Blood Pressure Titration Tool Evaluation

The 9 nurses who participated in the evaluation of the blood pressure titration tool had a median age of 29 years (range 25-63 years), 3 years of ICU experience (range 1-19 years) and 3 participants were male. The training session lasted an average of 3 min (range 1.4-4.1 min) before the participant completed the assigned training tasks correctly.

Figure 4.8a shows that the median time to reach the target blood pressure was 4.1 min (range 1.6-7.8 min) when nurses used the advisory system and 10.2 min (range 4.1-26.8 min) when they did not. A repeated measures Friedman's ANOVA found the differences statistically significant ($p = 9.1 \cdot 10^{-6}$, $\chi^2 = 16.69$). Nurses made a median of 4 SNP infusion rate changes (range 1-7) before reaching the target blood pressure when the advisory system was used and 6 without the advisor (range 1-41), see Figure 4.8. A repeated measures Friedman's ANOVA found that these differences were significant ($p = 3.7 \cdot 10^{-3}$, $\chi^2 = 8.45$).

An overshoot in MAP of more than 10 mmHg below the target MAP occurred in three cases without the advisory system and in two cases with the advisory system available (Fisher's exact test $p = 1$). A MAP $< 70$ mmHg occurred during two cases where the advisory system was not available ($p = 0.49$).

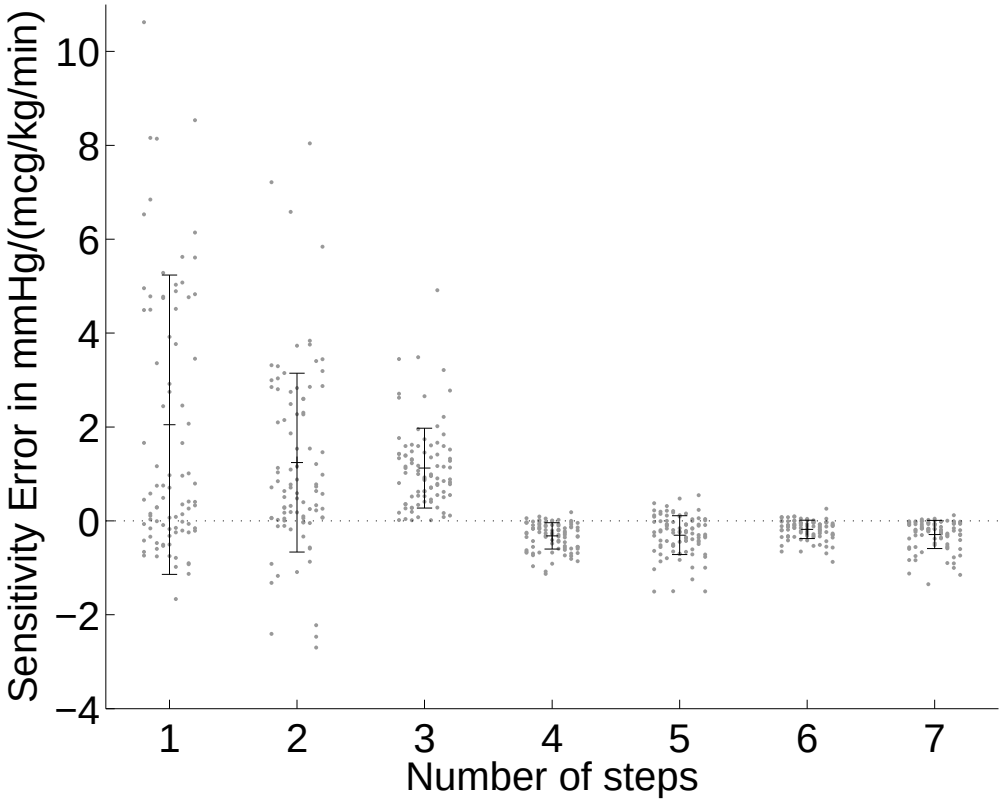


**Figure 4.6**: The error in our estimation of sensitivity to dopamine for 100 simulated patients. The error decreases rapidly with each of the first four step changes in the dopamine infusion rate, with each step lasting 15 min. The bars show the mean±standard deviation of the error at each step.

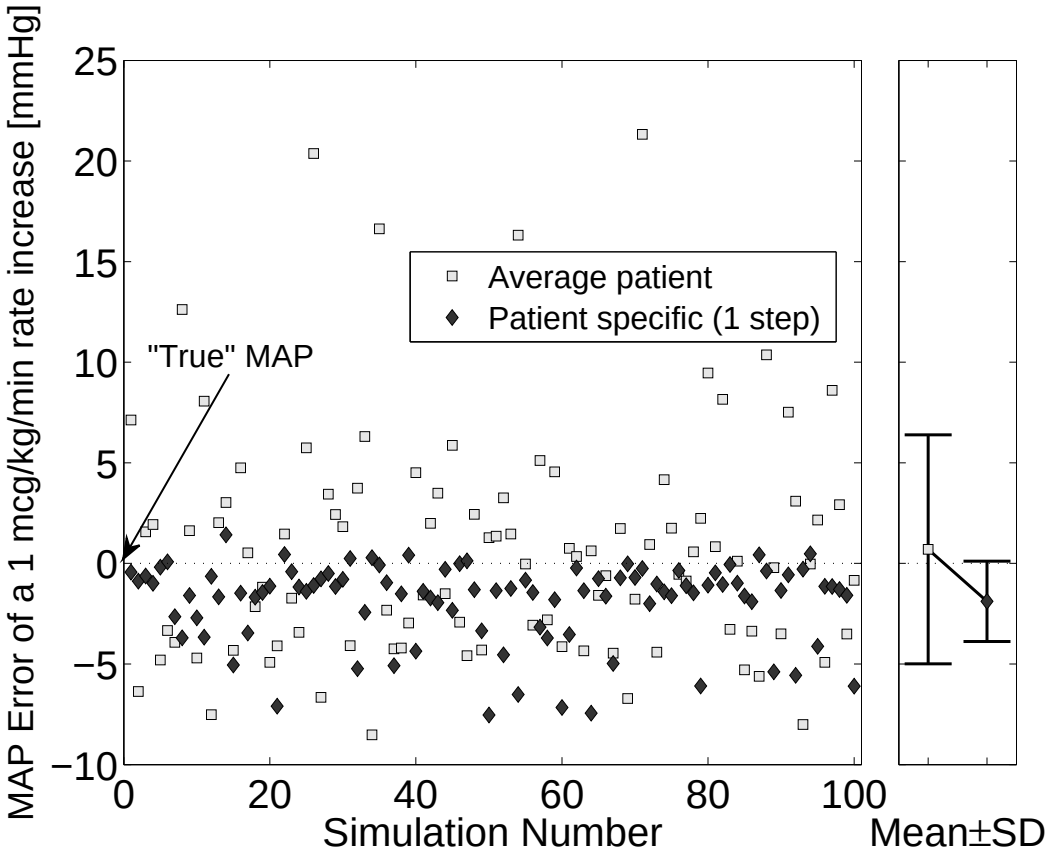


**Figure 4.7**: The error in our prediction of steady-state mean arterial blood pressure (MAP) after an increase in dobutamine infusion rate of $1\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$: The open boxes show the error when the prediction algorithm uses an "average" patient's sensitivity with $MAP_{max}$=28.6 mmHg and $I_{slope} = 2.11\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$. The black diamonds show the error after one infusion rate step change. It has a bias resulting in a mean underestimation of $I_{slope} = -0.30\,\mathrm{mcg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$ and $MAP_{max} = -3.75\,\mathrm{mmHg}$. The bars in the right panel show mean±standard deviation of the prediction error.

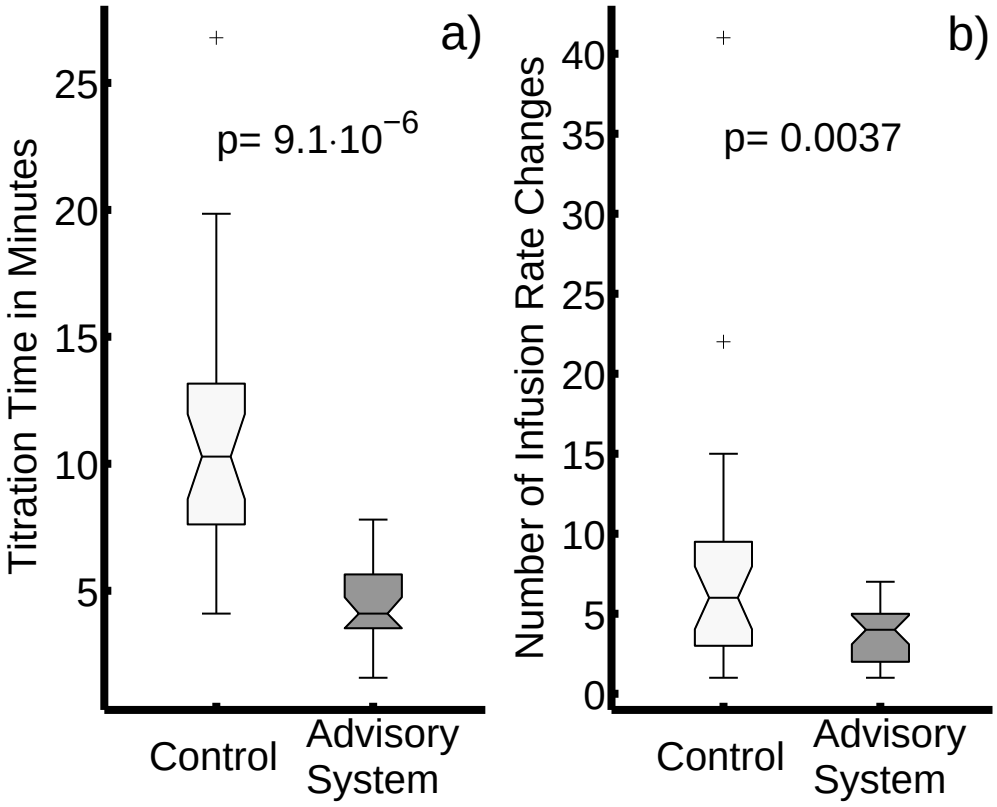


**Figure 4.8**: User performance with and without the advisory system: a) The gray box and whisker icon shows the time to reach a target blood pressure when using the advisory system. The white icon shows the time without the advisor. Each icon shows the lowest value, the lower quartile, the median value, the upper quartile and the uppermost value. A plus sign (+) shows outliers (>1.5x box heights above or below the box). b) The number of sodium-nitroprusside infusion rate changes required to reach the target pressure, using the same icons as panel a.

Figure 4.9 shows the participants' self assessment of the workload involved in bringing the blood pressure to the target with and without the advisory system. A Friedman's ANOVA found significant differences for all 6 workload scores. The support tool's greatest perceived benefit was a reduction in mental workload ($p = 3.1 \cdot 10^{-5}$), frustration ($p = 3.8 \cdot 10^{-5}$) and effort ($p = 5.1 \cdot 10^{-4}$).

## 4.5 Discussion

When nurses used an advisory system that predicted MAP 5 min into the future, they reached a target blood pressures in significantly less time (4.1 vs. 10.2 min), they made fewer changes in the SNP infusion rate (4 vs. 6) and avoided inducing hypotension (MAP<70 mmHg). The clinical use of the advisory system in an ICU could reduce nursing workload and inadvertent hypotension, thereby potentially increasing patient safety. The decrease in self-reported mental workload and effort, for this rather challenging titration test, was statistically significant. Many nurses asked when the system would be available for clinical use.

Our simulations show that an individual patient's sensitivity to SNP, dopamine and dobutamine can be accurately identified as the drugs are being infused. When the patient's individual sensitivity is used to predict MAP, the steady-state prediction improved in accuracy by 75% for SNP, 82% for dopamine and 52% for dobutamine over the MAP predicted using the "average" patient's sensitivity.

Population based models for SNP[9, 14] dopamine[11, 14] and phenylephrine[15] exist. Such models have been show to work well in predicting anesthesia effects for the "average patient".[16, 17] However, the interpatient variability in response to a drug can vary significantly. Given a $1\,\mathrm{mcg} \cdot \mathrm{kg}^{-1} \cdot \mathrm{min}^{-1}$ infusion of SNP will decreasing a patient's MAP anywhere from 6.5-171 mmHg (20 mmHg for the average patient), knowing the patient's specific sensitivity is important when predicting future MAP values.

Our results add to previously reported work in two ways:

1. A new target blood pressure can be reached quicker and with fewer infusion rate changes when nurses use an advisory system that identifies the patient's sensitivity to SNP and displays the predicted MAP 5 min in the future.

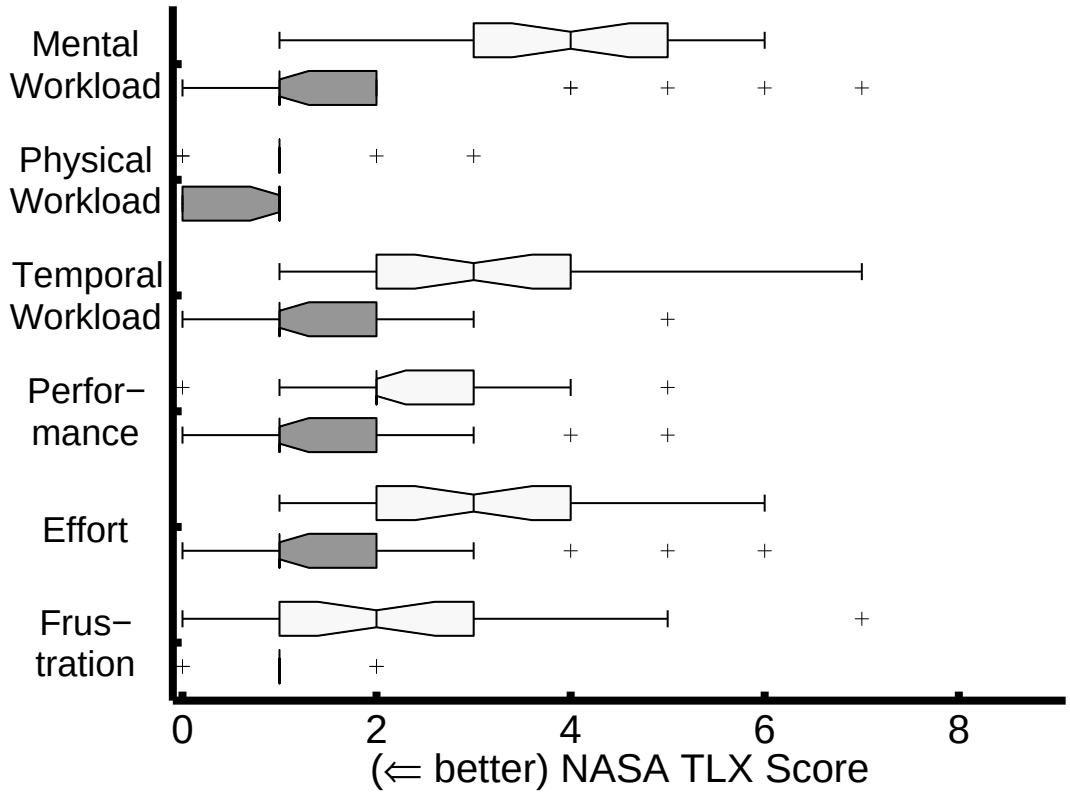


**Figure 4.9**: NASA TLX self-reported workload scores with and without use of the advisory system. Dark gray indicates the availability of the advisory system. Each icon shows the lowest value, the lower quartile, the median value, the upper quartile and the uppermost value. A plus sign (+) shows outliers (>1.5x box heights above or below the box). All decreases in workload were statistically significant.

2. The adaptive filter approach, using the MAP response to a small step change in the drug infusion rate, can identify the patient's sensitivity to SNP, dopamine and dobutamine as the drug is being delivered. Previous methods identify patient sensitivity using the large initial step change that occurs as the drug infusion starts, or following a bolus injection.[18]

We expect that the performance improvements nurses demonstrated when titrating SNP with our advisory system can be generalized to dopamine and dobutamine. As seen in Appendices B and C it might work better for dopamine as the dobutamine identification was less accurate, while the dopamine identification also took much longer to produce an initial result (45 min instead of 10 min for dobutamine).

### 4.5.1 Existing Sensitivity Identification Methods

Interpatient variability in sensitivity is a significant problem when titrating drug delivery to achieve a clinical effect.[9] Two alternatives to nurses or physicians manually titrating blood pressure exist: Closed loop feedback controllers,[9, 14, 19–24] which automatically measure blood pressure and change infusion rates based on observed responses, and open loop advisory systems, which predict future blood pressure changes based on infusion rate changes but requires a clinician to adapt infusion rates. Sheppard[25–27] developed a closed loop controller that automatically adjusted a SNP infusion to control MAP in patients after cardiac surgery. In his animal research, Sheppard imposed pseudo-binary random changes in the SNP infusion rate to identify the animal's sensitivity but did not have the computing power or algorithms to implement this method in real time. Later, IVAC developed the SNP Titrator,[18, 22] which started by giving SNP at a slow infusion rate and measured the resulting change in blood pressure to obtain an estimate of the patient's sensitivity. Jaklitsch and Westenskow's vecuronium controller[28] used two test doses during the induction of neuromuscular blockage to measure patient sensitivity. While closed-loop feedback controllers are faster, provide more precise MAP control, and offer a reduction in nursing workload, there are risks associated with artifacts in the blood pressure transducer, the loss of nursing staff supervision. In addition to the risks, there are regulatory and product liability hurdles to be overcome.

In the current study we used a small step change in infusion rate that caused MAP to change by ±3 mmHg to estimate the patient's sensitivity. One might argue that using small step changes, which wait until steady state is reached, is superior to pseudo-binary random changes, with a wide range of frequencies (durations) and magnitude, as the step change method is less sensitive to noise, clinically more acceptable as the resulting MAP changes are small and consistent, and potentially easier to implement in a medical product. We also felt that the requirement for the clinician to make infusion rate changes, which ensures checking of recommended changes for plausibility and direct supervision by a clinician, reduces the risk associated with using the system and therefore prefer this method.

### 4.5.2 Limitations

The main limitation of our identification algorithm is that takes at least 15 min before a patient's sensitivity can been identified with reasonable accuracy. If a blood pressure change of ≥6 mmHg is acceptable, an estimate can be obtained in 3 min for SNP (see Figure 4.10) or approximately 10 min for Dopamine. The waiting period would not be a problem if the patient has already spend time in the ICU and our sensitivity identification could have been performed continuously using historical data from the immediate past. In this case a new targeted MAP can be achieved and maintained without waiting for the sensitivity estimation.

Another limitation of this study is that all data were collected using simulations. However, the SNP model[9] has been carefully validated using observations in humans,[9, 18] and the dopamine and dobutamine models have been validated in animals.[11, 12] Additionally, both animal models show the same drug effect behavior (linear and exponential saturating dose effect relationships respectively) as observed in patients with congestive heart failure.[29] While animal experiments could be performed to evaluate the effectiveness of the proposed method, an evaluation with patients could also be performed, as long as a vigilant clinician uses their best judgment about the suggested infusion rates and the infusion rate changes required to identify the patient's sensitivity are performed by hand and not automatically.

Additionally, vasopressors are not utilized purely to increase blood pressure, but are also used for inotropic and/or chronotropic support for the failing myocardium.

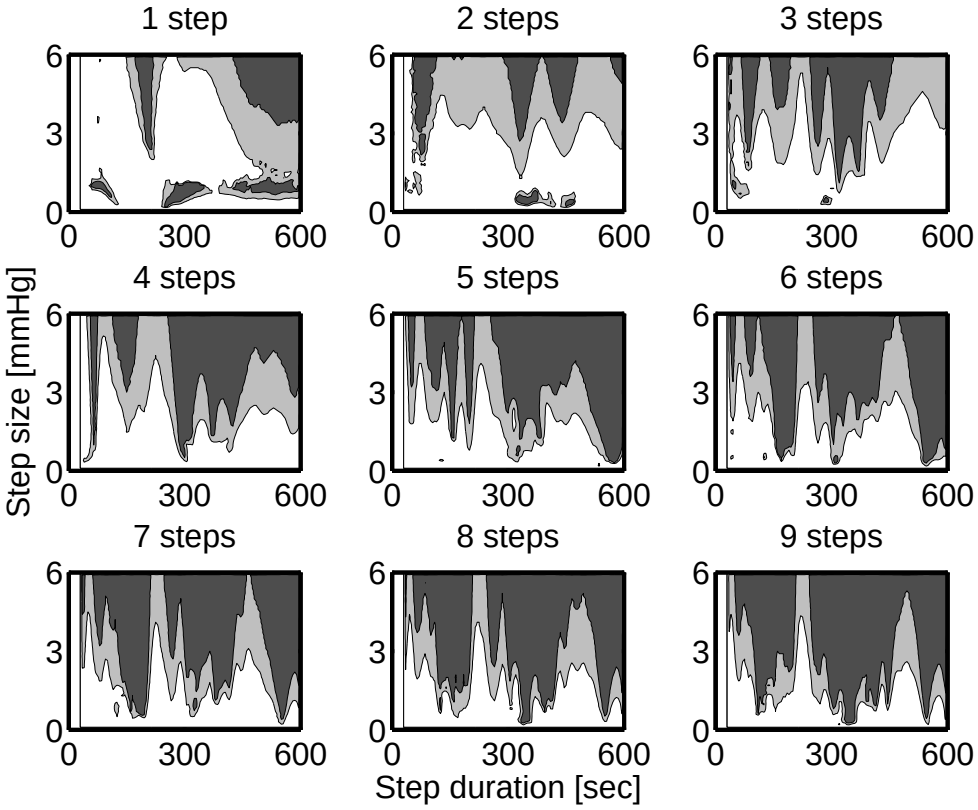


**Figure 4.10**: Sodium-nitroprusside (SNP) sensitivity identification error over most of the SNP model's parameter space. The error decreases with more steps, larger step magnitude and longer steps. We suggest using three steps, 3 mmHg in amplitude and 5 min in duration to identify patient sensitivity. Light gray areas indicate an error of less than $4\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$, and dark gray areas indicate an error of less than $2\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$, while white areas indicate an error of more than $4\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$. For readability the plot shows averaged values, where each point's eight neighbors were taken into consideration.

Thus, the titration algorithm would be much more complex for a drug like dobutamine, compared to vasopressin. By challenging the system using small step changes in vasopressor infusion rate one could not only identify the patient's MAP sensitivity but also the patients' heart rate (HR) sensitivity. If both are available advice could be based on a combination of clinician determined HR and MAP limits and with the suggestion to switch to different medications if the expected effect with the current medication could not reach the desired MAP goal.

We did not model norepinephrine, a drug commonly used in ICUs to increase MAP, as norepinephrine population models are hard[30] or impossible[31] to obtain. However, we feel that the dopamine sensitivity identification method (with a saturating dose response) might be the solution for this problem, which has yet to be explored using a swine model followed by observations in patients.

We assessed self-reported workload in simulations running 3.6 times faster-than-real time. This artificially high simulation speed is likely going to alter the absolute workload scores. However, as both the experimental condition and the control condition were performed at the same speed, the results of the comparison (relative changes) should still be valid. One might argue that a realistic distracter task could have been used instead; however we feel that by performing the study in the nurses break room, with realistic ambient noise and frequent alarm noises, as well as the fact that nurses certainly kept their patients' well-being and problems in mind during the experiment, nurses were already sufficiently distracted.

Finally, our approach needs a bidirectional interface to the infusion pump to change the rate, an arterial blood pressure transducer and an interface to the patient monitor to obtain MAP. It is reasonable to assume that in the near future infusion pumps will be connected directly to clinical information systems, where our advisory system will reside.

### 4.5.3 Conclusions

When ICU nurses used our advisory system, which predicts MAP 5 min in the future, they titrated blood pressure to the target faster with fewer infusion rate changes. Our patient-specific sensitivity estimation algorithm improved predictions of MAP during SNP, dopamine, and dobutamine infusions. Future work is needed to

implement and evaluate our approach in patients and verify the observed reduction of nursing workload, improved MAP control and shorter vasoactive drug titration times in a clinical evaluation.

## 4.6 Acknowledgements

The authors would like to thank all participating nurses at the University of Utah Hospital for their support. Additionally we would like to thank Boaz Markewitz, MD for clinical feedback on our titration experiment.

## 4.7 Appendix A: Identification of Optimal Step Size and Duration

We used Equation 4.2 to simulate the average patient's blood pressure response to SNP ($K_{SNP} = 20\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$), $MAP_{baseline}$=140 mmHg, $T_i$=30 sec, $T_c$=45 sec, $\beta$=0.2, $\tau$=40 sec). In order to test the system at steady state a typical infusion rate of $2\,\frac{\text{mcg}}{\text{kg}\cdot\text{min}}$ was administered to reach a goal MAP of 100 mmHg. As this extremely large blood pressure change would not be clinically acceptable, the initial blood pressure response was discarded. Next, the patient sensitivity identification algorithm described above was used to identify the average patient's sensitivity after simulating 1 to 9 step-increase/step-decrease iterations in the SNP infusion rate with MAP step sizes that ranged from 0.1-6.0 mmHg (0.1 mmHg increments) and steps that ranged in duration from 30-600 sec (1 sec increments). These simulations included white noise (10 mmHg prefiltered amplitude) and respiratory induced oscillations (3 mmHg amplitude with a frequency that ranged from 6-12/min).

Figure 4.10 shows the accuracy of the sensitivity estimation as a function of step size, step duration and number of steps. The top right subplot shows that the error was consistently below ($4\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$), for steps ranging from 2.8-3.5 mmHg in amplitude and 250-350 sec in duration, and below $2\,\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ for 310-350 sec duration at the same step amplitude. Therefore, we chose to use three steps, 3 mmHg in amplitude and 5 min in duration to identify patient sensitivity. However, a much quicker identification with bigger steps, e.g. using two 5 mmHg steps for 70 sec each it is also possible (Figure 4.10 top center subplot).

## 4.8 Appendix B: Dopamine Sensitivity Identification

Dopamine has a MAP response to a change in infusion rate that is linear over the clinical relevant range, just like SNP. Therefore the same identification algorithm can be used. Therefore its evaluation was performed parallel to the procedure for SNP.

### 4.8.1 Methods

For dopamine we used the model from Gingrich and Roy[11] to predict $\Delta MAP(s)$ when dopamine is given by intravenous infusion ($I_{Dopa}(s)$):

$$\Delta MAP = \frac{K_{Dopa} \cdot e^{-T_i \cdot s} \cdot I_{Dopa}(s) - MAP_{offset}}{1 + \tau \cdot s} \tag{4.4}$$

for $I_{Dopa} \geq MAP_{offset}/K_{Dopa}$. The model parameters are listed in Table 4.2.

### 4.8.2 Evaluation of Sensitivity Identification Performance

We simulated 100 patients and measured the accuracy of the linear sensitivity estimation methods described above. A baseline MAP $MAP_{baseline}(k)$ with a mean and SD of 74±11.5 mmHg, 20±2mmHg lower than the target MAP in the SNP experiment was used. The patient sensitivity estimation algorithm used 3 mmHg steps 15 min in duration and a filter strength $\alpha$ that ranged from 0.001 to 0.999, to allow for the slower drug response. Fifteen min were required after each step change in infusion rate, which is shorter than 25 min required ensuring steady state, but still allowed sufficient MAP change to occur. Additionally, one change to step 4 of the proposed identification algorithm was required: Negative MAP changes were not considered and instead replaced by the measured MAP. This prevents predicting unrealistic MAP reductions for dopamine infusion rates below the activation threshold ($MAP_{offset}/K_{Dopa}$).

### 4.8.3 Results

Figure 4.6 shows the difference between the algorithm's estimation of the simulated patient's sensitivity and the simulated patient's actual sensitivity to dopamine. After three 15 min steps the mean difference was $1.12 \pm 0.85 \, \frac{\text{mmHg}}{\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}}$. Again there is a trade-off between identification time and estimation error as a 1hr identification with

**Table 4.2**: Dopamine model parameters

| Parameter | Description | Symbol | Mean±Standard Deviation |
|---|---|---|---|
| $P_1$ | Patient's sensitivity | $K_{Dopa}$ | $12.3 \pm 10.09 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ (Limited to $K_{Dopa} > 0$) |
| $P_2$ | Minimum MAP increase, used to calculate the activation threshold. | $MAP_{offset}$ | $24.3 \pm 20.18\,\text{mmHg}$ (Always using the same SD value as $2 \cdot K_{Dopa}$) |
| $P_3$ | System time constant | $\tau$ | $5.74 \pm 1\,\text{min}$ |
| $P_4$ | Infusion delay | $T_i$ | $40 \pm 10\,\text{sec}$ (Same as for SNP) |
| $P_5$ | Baseline MAP | $MAP_{baseline}$ | $MAP_{goal} - 20 \pm 2\,\text{mmHg}$ |
| $P_6$ | Target MAP | $MAP_{goal}$ | $90 \pm 12.5\,\text{mmHg}$ (>70 mmHg) (Same as for SNP) |

Parameters used in the Dopamine transfer function (Equation 4.4). Values for the parameters $P_1 - P_3$ were obtained from Gingrich and Roy,[11] $P_4$ and $P_6$ reuses the value introduced for the SNP model and $P_5$ was picked to be 20±2 mmHg below the goal MAP to simulate a MAP with the need to raise it.

an error of $-0.32 \pm 0.28 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ would also be possible. When 100 simulated patients received a dopamine infusion of $2\,\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, the error in the steady-state MAP prediction improved from 2.8±16.0 mmHg to 2.2±1.7 mmHg because of the sensitivity identification (not shown). The estimation of the patient's sensitivity to dopamine had less patient-to-patient variability. The reason may be that the SNP model included recirculation, while the dopamine model did not. Recirculation, while physiologically true as the drug is not completely metabolized immediately, is only important when transient changes of MAP are of interest. However, for determining a patient's sensitivity to a drug only the steady state changes are relevant, for which the presence or absence of recirculation has no effect.

## 4.9 Appendix C: Dobutamine Sensitivity Identification

For Dobutamine, the MAP response to a change in infusion rate is nonlinear: For high infusion rates an increase in infusion rate causes a smaller change in MAP than it does for the same change at a lower infusion rate, until it finally completely disappears for very high infusion rates. Therefore, a different method to identify nonlinear sensitivities has to be used.

Otherwise this evaluation was performed parallel to the procedure for SNP.

### 4.9.1 Methods

For Dobutamine we propose a new exponential saturating dose response model, based on observations in canines by Kamijo et al.[12] It predicts the change in a patient's MAP ($\Delta MAP(s)$) when dobutamine is given by intravenous infusion $I_{Dobu}(s)$ as:

$$\Delta MAP = \frac{MAP_{max} \cdot \left(1 - exp\left(-I_{Dobu}(s) \cdot e^{-T_i \cdot s}/I_{slope}\right)\right)}{1 + \tau \cdot s} \tag{4.5}$$

The model parameters are listed in Table 4.3.

### 4.9.2 Exponential Saturating Sensitivity Identification

The following algorithm was used to identify the patient's sensitivity for Dobutamine, because of its nonlinear sensitivity curve. It could potentially be expanded to Norepinephrine, which demonstrates a similar behavior:

**Table 4.3**: Dobutamine model parameters

| Parameter | Description | Symbol | Mean±Standard Deviation |
|---|---|---|---|
| $P_1$ | Infusion rate contant | $I_{slope}$ | $2.1 \pm 0.5 \frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ (Limited to $I_{slope} > 0$) |
| $P_2$ | Maximum MAP increase | $MAP_{max}$ | $28.6 \pm 11$ mmHg (Limited to $MAP_{max} > 0$) |
| $P_3$ | Infusion delay | $T_i$ | $40 \pm 10$ sec (Same as for SNP) |
| $P_4$ | System time constant | $\tau$ | $120 \pm 30$ sec |
| $P_5$ | Baseline MAP | $MAP_{baseline}$ | $90 \pm 14.5$ mmHg (>70 mmHg) (Same as for Dopamine) |

Parameters used in the Dobutamine transfer function (Equation 4.5). Values for the parameters $P_1$ and $P_2$ were obtained by fitting an exponential function to Kamijo et al.'s[12] response curve observed in canines. The best fit had a RMSE of 0.932 mmHg. For $P_4$ we made a reasonable assumption using the Dobutamine package insert and for the $P_5$ we reused the values introduced for the Dopamine model.

1. Begin with an initial infusion transport delay ($T_i$) of 10 sec and initial values for $MAP_{max}$ of 25 $\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$ and $I_{slope}$ of $2\,\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$.

2. Shift the original infusion rate $I_{Drug}(t)$ forward in time using $I_{shifted} = I_{Drug}(t - T_i)$ so that it is more closely aligned in time with the corresponding changes in $MAP_{measured}(t)$.

3. Predict the patient's MAP during a drug infusion using $MAP_{predicted} = MAP_{baseline} + MAP_{max} \cdot (1 - exp\,(-I_{shifted}(t)/I_{slope}))$ and calculate the root mean square error (RMSE) between $MAP_{predicted}(t)$ and $MAP_{measured}(t)$. For our work $MAP_{baseline}$ was known from the initial MAP measurements before the drug infusion began, but it could also become a parameter optimized for.

4. Select new values for $T_i$, within the limits of 0-100 sec, $MAP_{max}$, within the limits of $1 - 100$ $\frac{\text{mmHg}}{\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}}$, and $I_{slope}$, within the limits of $0.5 - 3.5\,\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, and repeat steps 2 and 3 in a Hessian multiparameter optimization, using MATLAB's fmincon with changes between $10^{-8} - 0.1$, to find a combination of these three parameters that results in the lowest RMSE.

5. When the optimization RMSE does not improve anymore ($< 10^{-8}$ to previous value) or the maximum number of iterations (250) is reached the identified sensitivity values ($MAP_{max}$ and $I_{slope}$) and the infusion delay ($T_i$) are reported. The obtained $MAP_{max}$, $I_{slope}$ and $T_i$ can now be used to predict $MAP(t)$ 15 min into the future. To better reflect the transient changes in the MAP prediction, the infusion rate could be filtered using the low-pass filter described in step 3 of the linear sensitivity identification method. A filter strength $\alpha$=0.0085, corresponding to a system time constant $\tau$=120 sec, would be a good initial choice.

### 4.9.3 Evaluation of Sensitivity Identification Performance

We simulated 100 patients and measured the accuracy of the sensitivity estimation. For dobutamine, the exponential saturating sensitivity estimation algorithm was used with infusion rate steps of $\pm 0.5\,\text{mcg}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ 10 min in duration. Maximum

MAP $MAP_{max}(k)$ was selected using Equation 4.3 with a mean and SD of 28.6±1.1 mmHg, the infusion rate constant $I_{slope}(k)$ with $2.1 \pm 0.5\,\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, the system time constant $\tau(k)$ with 120±30 sec, and the baseline $MAP(k)$ the same as for dopamine.

### 4.9.4 Results

When 100 simulated patients received a dobutamine infusion, and their sensitivity was estimated by making a single $+1\,\text{mcg} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ change of infusion rate, Figure 4.7 shows that the estimation of MAP was in error by -1.88±1.99 mmHg (range -7.54... 1.41 mmHg). The error did not change significantly with more steps; the third step error was -1.91±1.93 mmHg (range -7.84... 1.43 mmHg). Without sensitivity estimation the predicted MAP was in error by 0.70±5.69 mmHg (range -8.52... 21.33 mmHg).

## 4.10 References


1. Varon J. Treatment of acute severe hypertension: current and newer agents. Drugs. 2008;68(3):283–297.

2. Müllner M, Urbanek B, Havel C, Losert H, Waechter F, Gamper G. Vasopressors for shock. Cochrane Database Syst Rev. 2004;(3):CD003709.

3. Hollenberg SM. Vasopressor support in septic shock. Chest. 2007 Nov;132(5):1678–1687.

4. Overgaard CB, Dzavik V. Inotropes and vasopressors: review of physiology and clinical use in cardiovascular disease. Circulation. 2008 Sep;118(10):1047–1056.

5. Syroid ND, Agutter J, Drews FA, Westenskow DR, Albert RW, Bermudez JC, et al. Development and evaluation of a graphical anesthesia drug display. Anesthesiology. 2002 Mar;96(3):565–575.

6. Drews FA, Syroid N, Agutter J, Strayer DL, Westenskow DR. Drug delivery as control task: improving performance in a common anesthetic task. Hum Factors. 2006 Spring;48(1):85–94.

7. Stokes A, Wickens C. Aviation Displays. In: Wiener E, Nagel D, editors. Human Factors in Aviation. San Diego, CA: Academic Press; 1988. p. 387–431.

8. DeLucia P, Ott T, Palmieri P. Performance in Nursing. In: Durso F, editor. Reviews of Human Factors and Ergonomics. vol. 5. Santa Monica, CA: Human Factors and Ergonomics Society; 2009. p. 1–40.

9. Slate JB. Model-based design of a controller for infusing sodium nitroprusside during postsurgical hypertension [PhD Thesis]. University of Wisconsin-Madison. Madison, WI; 1980.

10. Devlin JW, Seta ML, Kanji S, Somerville AL. Fenoldopam versus nitroprusside for the treatment of hypertensive emergency. Ann Pharmacother. 2004 May;38(5):755–759.

11. Gingrich KJ, Roy RJ. Modeling the hemodynamic response to dopamine in acute heart failure. IEEE Trans Biomed Eng. 1991 Mar;38(3):267–272.

12. Kamijo T, Tomaru T, Miwa AY, Nakamura F, Kido H, Sugimoto T, et al. The effects of dobutamine, propranolol and nitroglycerin on an experimental canine model of congestive heart failure. Jpn J Pharmacol. 1994 Jul;65(3):223–231.

13. Hart S, Staveland L. Development of NASA-TLX (Task Load Index) Results of Empirical and Theoretical Research. In: Hancock P, Meshkati N, editors. Human Mental Workload. Amsterdam: North Holland Press; 1988. p. 139–183.

14. Yu C, Roy RJ, Kaufman H, Bequette BW. Multiple-model adaptive predictive control of mean arterial pressure and cardiac output. IEEE Trans Biomed Eng. 1992 Aug;39(8):765–778.

15. Huang JW, Roy RJ. Multiple-drug hemodynamic control using fuzzy decision theory. IEEE Trans Biomed Eng. 1998 Feb;45(2):213–228.

16. Kern SE, Stanski DR. Pharmacokinetics and pharmacodynamics of intravenously administered anesthetic drugs: concepts and lessons for drug development. Clin Pharmacol Ther. 2008 Jul;84(1):153–157.

17. Olofsen E, Dahan A. Population pharmacokinetics/pharmacodynamics of anesthetics. AAPS J. 2005;7(2):383–389.

18. Bednarski P, Siclari F, Voigt A, Demertzis S, Lau G. Use of a computerized closed-loop sodium nitroprusside titration system for antihypertensive treatment after open heart surgery. Crit Care Med. 1990 Oct;18(10):1061–1065.

19. Martin JF, Schneider AM, Quinn ML, Smith NT. Improved safety and efficacy in adaptive control of arterial blood pressure through the use of a supervisor. IEEE Trans Biomed Eng. 1992 Apr;39(4):381–388.

20. Martin JF, Smith NT, Quinn ML, Schneider AM. Supervisory adaptive control of arterial pressure during cardiac surgery. IEEE Trans Biomed Eng. 1992 Apr;39(4):389–393.

21. Chitwood WRJ, Cosgrove DMr, Lust RM. Multicenter trial of automated nitroprusside infusion for postoperative hypertension. Titrator Multicenter Study Group. Ann Thorac Surg. 1992 Sep;54(3):517–522.

22. Westenskow DR, Meline L, Pace NL. Controlled hypotension with sodium nitroprusside: anesthesiologist versus computer. J Clin Monit. 1987 Apr;3(2):80–86.

23. Jastremski M, Jastremski C, Shepherd M, Friedman V, Porembka D, Smith R, et al. A model for technology assessment as applied to closed loop infusion systems. Technology Assessment Task Force of the Society of Critical Care Medicine. Crit Care Med. 1995 Oct;23(10):1745–1755.

24. Crouse L. Advances in Pharmacological Cardiac Stress Testing: Arbutamine Closed-Loop Feedback System. Echocardiography. 1998 May;15(4):393–400.

25. Sheppard L, Sayers BM. Dynamic analysis of the blood pressure response to hypotensive agents, studied in postoperative cardiac surgical patients. Comput Biomed Res. 1977 Jun;10(3):237–246.

26. Sheppard LC. The computer in the care of critically ill patients. Proceedings of the IEEE. 1979 Sep;67(9):1300–1306.

27. Sheppard LC. Computer control of the infusion of vasoactive drugs. Ann Biomed Eng. 1980;8(4-6):431–434.

28. Jaklitsch RR, Westenskow DR. A model-based self-adjusting two-phase controller for vecuronium-induced muscle relaxation during anesthesia. IEEE Trans Biomed Eng. 1987 Aug;34(8):583–594.

29. Leier CV, Heban PT, Huss P, Bush CA, Lewis RP. Comparative systemic and regional hemodynamic effects of dopamine and dobutamine in patients with cardiomyopathic heart failure. Circulation. 1978 Sep;58(3 Pt 1):466–475.

30. Beloeil H, Mazoit JX, Benhamou D, Duranteau J. Norepinephrine kinetics and dynamics in septic shock and trauma patients. Br J Anaesth. 2005 Dec;95(6):782–788.

31. Johnston AJ, Steiner LA, O'Connell M, Chatfield DA, Gupta AK, Menon DK. Pharmacokinetics and pharmacodynamics of dopamine and norepinephrine in critically ill head-injured patients. Intensive Care Med. 2004 Jan;30(1):45–50.

# CHAPTER 5

# A FAR-VIEW INTENSIVE CARE UNIT MONITORING DISPLAY ENABLES FASTER TRIAGE*

## 5.1 Abstract

Nurses perform the majority of the clinical tasks in an intensive care unit. However, current patient monitors were not designed to support a nurse's monitoring and work flow task. Nurses constantly triage patients, deciding which patient is currently in most need of care. This includes integration of a patient's vital signs with therapeutic device information from multiple sources. To obtain this information they often have to enter the patient's room.

This study addresses three hypotheses: Information provided by far-view displays a) reduces the amount of time that it takes to determine which patient needs care first, b) increases the accuracy of assigning priority to the right patient and c) reduces nurses mental workload.

We developed two far-view monitoring displays to be read from a distance of 3-5 meters. They display numeric values for five vital signs, trends and alarms, infusion pump status, and therapy support indicators for blood oxygen saturation and minute ventilation. To evaluate the displays nurses were asked to use the displays to decide which of two patients required their attention first, making 20 decisions with each display (Control: Patient monitor next to an infusion pump).

Sixteen nurses (median age of 27.5 years with 2.75 years of experience) participated in the study. Using the two far-view displays, nurses more accurately and rapidly

*Submitted to Nursing Research in June 2010: Görges M, Kück K, Koch SH, Agutter J, Westenskow DR. A Far-view Intensive Care Unit Monitoring Display Enables Faster Triage. Nurs Res. 2010 

identified stable patients, and syringe pumps that were nearly empty. Median decision times were 11.3 sec for the Bar display, 12.4 sec for the Clock display, and 17.2 sec for the Control display.

Far-view displays reduced median decision making times by 4.8-5.9 sec, increased accuracy in assignment of priority in 2/7 scenarios, and reduced nurses' frustration with the triaging task. In a clinical setting the proposed far-view display might reduce nurses' workload and thereby increase patient safety.

## 5.2 Introduction

### 5.2.1 Background

Nurses perform the majority of the clinical bedside tasks in an intensive care unit, with the most frequent tasks being manually administering drugs or changing drug infusion rates (2.3/hr per patient room), silencing an alarm (1.3/hr), charting (1.1/hr), and performing patient assessments (0.7/hr).[1] However, current bedside patient monitors are not specifically designed to support the most common nursing tasks and work flow.

### 5.2.2 Problems with Current Monitoring

The general presentation of information is the same for all clinicians, even though nurses have a unique set of tasks that are different from the other clinicians'. The reason for this is that the first patient monitors were oscilloscopes that displayed electrocardiogram (ECG) and blood pressure (BP) waveforms. Later LED displays showing numeric values and alarms were added to support vigilance and medical decision making. Today LCD displays show essentially the same information in the same format, still following the single-sensor, single indicator paradigm, showing one waveform and/or numeric for each sensor.[2]

In addition, data from infusion pumps and ventilators are not integrated into the bedside display, even though this has been shown to enhance decision making.[3, 4] When information integration is performed, this is to facilitate documentation of vital sign in the electronic medical record.

### 5.2.3 Prioritizing Attention to Patients

Triaging patients, deciding which patient is currently in most need of nursing care, is a frequently performed task when nurses care for more than one patient. In addition to triaging patients, who they are assigned to and familiar with, nurses frequently need to triage unfamiliar patients, e.g., when covering for their co-workers who have left the unit for transports.[5, 6]

Triaging could be facilitated by integration or consolidation of disparate information from multiple bedside devices, such as the vital signs patient monitor, the ventilator, the infusion or syringe pumps and the continuous cardiac output monitor into one display highlighting the patient's need for attention. Ideally an integrated display allows at-a-glance assessments of the patient's condition without requiring the clinician entering the patient's room. The new display is not intended to replace the traditional display, rather augment it by providing summary information when a provider is not at the bedside. A monitor displaying our far-view display might revert to displaying a traditional waveform display when a health care provider is present at the bedside. Alternatively, the far-view display might be shown on a small LCD display by the doorway to the patient's room or on a small handheld mobile device, like a cell phone.

### 5.2.4 Purpose of the Study

This study addresses the following three hypotheses: 1) the information provided by far-view displays allow nurses to rapidly identify which of two patients needs immediate attention most, 2) the accuracy of clinical decisions and assessments will be increased with the consolidation of patient- and device information and 3) far-view displays will reduce the nurses' mental workload.

## 5.3 Methods

### 5.3.1 Far-View Display Development

We developed the far-view displays: the Bar display in Figure 5.1 and the Clock display in Figure 5.2, using a user centered design process[7] with an interdisciplinary team of nurses, physicians, designers, engineers, and human factors experts. The displays show a) numeric values and trends for heart rate (HR), blood oxygen satura-

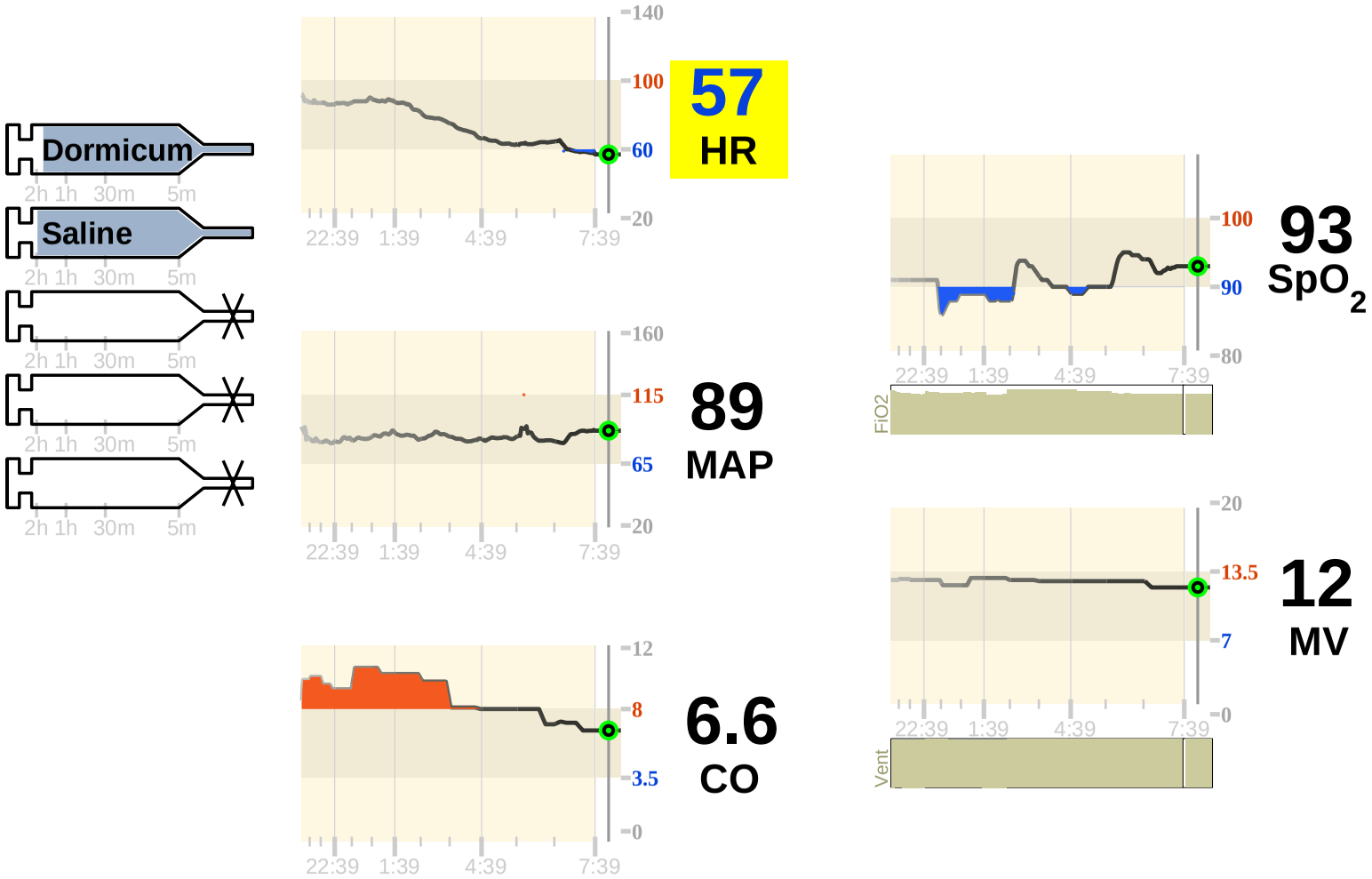


**Figure 5.1**: Far-view display in "Bar" presentation, showing a linear 12 hr trend looking like a strip chart. The display is indicating a heart-rate (HR) alarm, at a HR of 57 bpm that continuously decreased over the last 6 hr, in a patient that previously had low blood oxygen saturation (SpO2) and high cardiac output (CO) values. Two syringe pumps are currently running, with the X over the tip of the syringe pump indicating a stopped infusion.

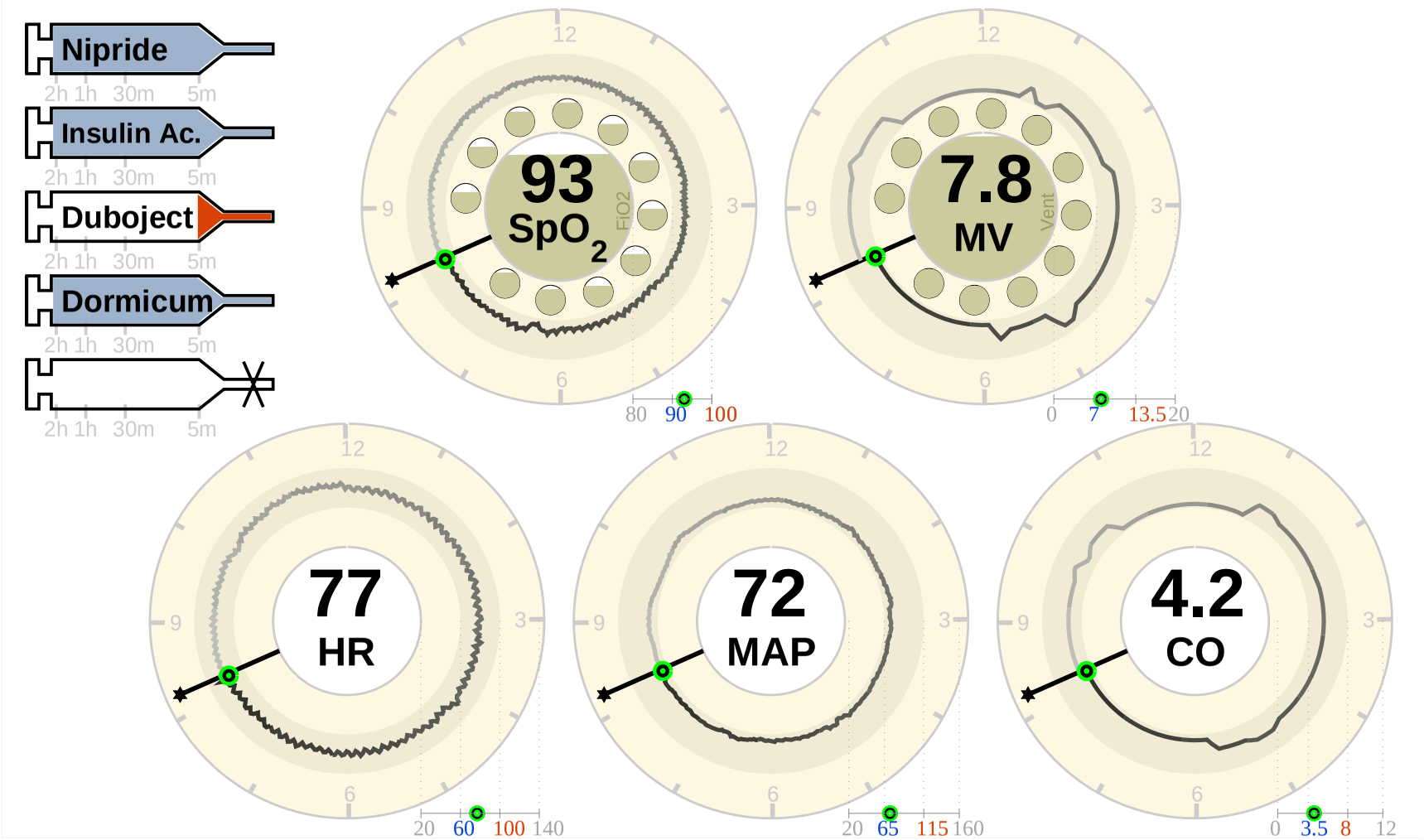


**Figure 5.2**: Far-view display in "Clock" presentation, showing the trend information on a circle looking like a 12 hr clock. The display is presenting a patient with stable vital signs that had a slightly reduced, but not alarming, mean arterial pressure (MAP) and has four currently running medications. The third syringe pump ("Dobuject") is going to be empty in approximately 9 min.

tion (SpO2), continuous cardiac output (CCO), mean arterial pressure (MAP), and ventilation minute volume (MV), b) infusion or syringe pump information, c) alarm status, and d) therapy support indicators for SpO2, MV, and MAP.

#### 5.3.1.1 Trend Component

Both displays show five 12 hr trends, sampled in 2 min intervals and median averaged over a 10 min window: The clock-like display (Clock), presents the trend information on a circle with a time scale of 12 hr (Figure 5.2). The strip-chart display (Bar), presents the trend information as a horizontal line, which collapses trends into 1 hr linear sections as they get older (Figure 5.1).

#### 5.3.1.2 Alarm Indicator

The normal, nonalarming areas are indicated with a darker background color. When a variable crosses an upper alarm limit the area between the alarm threshold and the measured value is filled in red. When a variable crosses below the lower alarm limit the area between the alarm threshold and the measured value is filled in blue. Alarms are indicated by highlighting the out of range numeric variable in yellow color, which could flash in an actual implementation.

#### 5.3.1.3 Syringe Pump Information

The syringe pump icon shows the name of the medication and the time until the bag (in case of a volumetric infusion pump) or syringe will be empty. Times until empty are displayed to remind nurses to get more medications from a storage location or order them from the pharmacy, as nurses were observed to use infusion pump alarms as reminders.[1] The infusion rate is indicated by the diameter of the syringe icon (high, medium and low).

#### 5.3.1.4 Therapy Support Indicator

A trend of FiO2 is shown alongside the SpO2 trend to indicate therapeutic support; Ventilator provided MV (MVmech) trend is shown alongside the measured MV trend. In the Bar display the support is indicated by filling of the box below the trend, in the Clock display with 11 filled circles in 1 hr intervals as well as the center circle for the current support. Therapy support for MAP, which predicted

MAP changes caused by vasoactive drug infusions,[8] was implemented but not tested in this evaluation.

### 5.3.2 Far-View Display Evaluation

To evaluate the far-view displays nurses were placed in front of a computer monitor showing two copies of Figures 5.1 and 5.2, one copy (size: 10" diagonal) for each of 2 patients. They were asked to use the displays to decide which of the 2 patients required their attention first, making 20 decisions using each of the 2 novel far-view displays. Additionally, the data from the same pairs of patients was shown on a control display (Figure 5.3): a cardiac patient monitor alongside an infusion pump, and the volunteers again made twenty decisions.

A 2-4 (scenario repetition) x 3 (displays) x 7 (scenarios) repeated-measures within-subject experimental design was used with two independent variables, the display presentation and the scenario presented, and one controlled variable, the scenario order.

#### 5.3.2.1 Scenarios

We selected 40 12-hr trend sections containing epochs of patient vital signs and drug infusion rate data from the IMPROVE data library[9] based on the criteria listed in Table 5.1. Each selected epoch had MV and CO values available for at least 10 hours of the 12 hr period and contained at least one change in minute ventilation. For all but 2 epochs, which were used to show near empty syringe pumps, we set the time remaining for active infusion pumps to 60-150 min. In case a patient received less than 3 medications we added a "Saline" infusion with a rate of 45-55 mL/hr. As CO was only measured intermittently, we displayed the last measured value until a new value was measured. As we wanted to keep the same alarm thresholds for all scenarios, we added a constant bias to some vital signs to achieve the desired presentation.

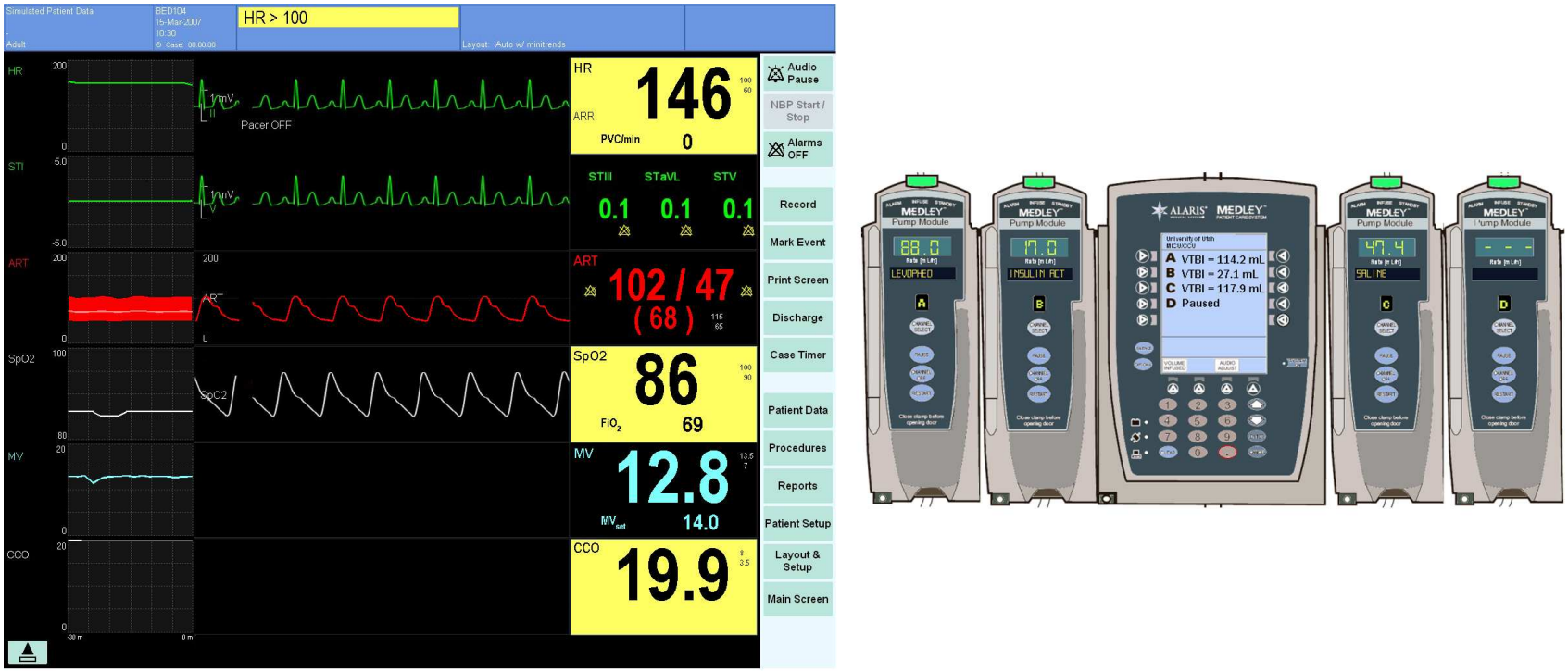


**Figure 5.3**: Control display consisting of a Dräger Kappa XLT patient monitor and an Alaris Medley infusion pump. The display is presenting a patient with multiple alarms: high heart-rate (HR), low blood oxygen saturation (SpO2) and high cardiac output (CCO). Three infusion pumps are currently running.

**Table 5.1**: Differences between critical and less critical patients

| Scenario name and description | Patient most in need of attention | Patient less in need of attention | Repetitions per display |
|---|---|---|---|
| "Vital Signs Stable" * Two stable patients with no alarms in the past 12 hr. | 0 alarms in the 12 hr trend | | 2 |
| "Past Alarms" * Two patients with multiple alarms in the past 12 hr but no current alarms. | 0 active alarms | | 3 |
| | 20+ min with MAP alarms and average MAP >50 mmHg OR 20+ min with HR alarms and average HR <115 bpm | Fewer, shorter and smaller deviations | |
| "One Alarm" § Two previously stable patients, of which one has a current HR or MAP alarm. | 0 HR and 0 MAP alarms in first 11 hr | | 4 |
| | 10+ min of HR alarms in last hr and active HR alarm OR 10+ min of MAP alarms in last hr and active MAP alarm | 0 active alarms in last hr | |
| "Multiple Alarms" * Two very sick patients, with multiple alarms, of which one has more active alarms than the other. | Average SpO2 <88% and average FiO2 >80% OR Average MAP <60 mmHg and average HR>110 bpm OR Average CO <3.5 L/min OR Average CO > 10 L/min and more than one vasopressor | | 4 |
| | 2+ active alarms | 1+ active alarms (always fewer than more critical patient) | |

**Table 5.1** continued

| Scenario name and description | Patient most in need of attention | Patient less in need of attention | Repetitions per display |
|---|---|---|---|
| "Ventilator Setting" † Two stable patient, of which one has a higher FiO2, more ventilator support, and a lower SpO2 than the other. | > 90% MVmech<br>88% < Average SpO2 < 91%<br>Average MV < 13.5 L/min<br>Average FiO2 > 50% | < 40% MVmech<br>Average SpO2 > 92% | 2 |
| "Pump Reminder" * Two stable patients of which one has an infusion running out in less than 17 min. | 0 alarms in the 12 hr trend | | 2 |
| | 5-10 min until syringe empty<br>OR<br>13-17 min until syringe empty | All syringes > 60 min until empty | |
| "Approaching Alarm" § Two patients, of which one has a HR, MAP or SpO2 trend heading towards the alarm threshold with an alarm occurring in the next 4min. | 0 HR and 0 MAP and 0 SpO2 alarms in first 11 hr of trend<br>0 active alarms | | 3 |
| | HR or MAP or SpO2 alarm will occur in 4 min<br>HR or MAP or SpO2 trend towards alarm threshold for 45+ min | No HR or MAP or SpO2 trend towards alarm threshold | |

* Both epochs could be from different patients in the dataset.
§ The same patient was used but the epoch of the less urgent patient started one hr earlier and ending one hr earlier than the epoch of the more urgent patient.
† Search restricted to two patients with varying levels of SpO2 and mechanical MV.

#### 5.3.2.2 Power Analysis

We assumed that when the new display is used 3 more of the 20 answers would be correct than with the control display. To achieve a Power of 0.8 for Fisher's exact test, with an $\alpha_{crit} = 0.05$, and 20 scenarios per display and participant, we found a sample size requirement of $n = 12$ participants.

#### 5.3.2.3 Participants

Approval from the University of Utah Health Sciences Center's institutional review board (ClinicalTrials.gov identifier: NCT00714012) was obtained. Sixteen medical intensive care unit nurses participated in the study.

#### 5.3.2.4 Training and Quiz

In a training session lasting 3.5-7 min we explained each element in the three displays (both far-view displays and our control display) using six PowerPoint slides (Microsoft Corporation, Redmond, WA). Examples when alarms were active and syringes were empty were demonstrated. Participants were encouraged to ask questions. After completion of the training session the participants were given four questions to test their ability to detect a MAP alarm, a SpO2/FiO2 support indicator difference, a HR trend towards an alarm state and 2 stable patients. Participants were shown the same information on all three displays placed side-by-side. If they could not answer all four questions correctly, they received additional training and repeated the quiz. Failure to complete the quiz the second time was an exclusion criterion.

#### 5.3.2.5 Apparatus

We compared the two far-view displays with a Kappa XLT patient monitor (Draeger Medical Inc, Telford, PA) and an Alaris Medley infusion pump (CareFusion, San Diego, CA). The exact same data were shown on the two far-view displays (Figures 5.1 and 5.2) and on the XLT monitor and Medley infusion pump (Figure 5.3). Table 5.1 describes the trend section (data epoch) shown on each display.

The far-view displays were created using MATLAB (The MathWorks Inc., Natick, MA). The control display used screenshots of the patient monitor and infusion pump, which were stripped of numbers, text and the MAP waveform and consequently filled

with the appropriate numbers, trends, medication names, and MAP waveforms using MATLAB. The experiment was performed using a web-based testing system that automatically measured decision time and accuracy.

#### 5.3.2.6 Scenario and Procedure

At the start of each scenario the participant was told: "Your co-worker had to leave the unit to help another nurse transfer a patient to the CT. She asked you to take care of her two patients while she is gone." and asked: "Please choose which of the two patients requires your attention first." In the event that the participant could not decide they could select: "Both patients are equally in need of attention."

Each participant was shown the data from all 20 scenarios, in all display conditions, one display at a time. The scenario order was randomized and we counterbalanced the order with the even numbered participant receiving the inverse order of the odd numbered participant. The display order was blocked and repeated with every 6 participants using a Latin-square design. The position of the correct answer was randomly assigned to be equally often on top and bottom. Participants were encouraged to make the right decision as fast as possible and spend 100% of their attention on the decision task.

We recorded the times to reach the decision and whether the answer was correct. After each block of questions were finished, participant completed the NASA Task Load Index (TLX) rating sheet.[10] After the conclusion of the study, participants were asked to select which of the three displays they liked best. The experiment was performed in the break room of the ICU, providing for realistic ambient noise and distractions.

#### 5.3.2.7 Data Analysis

We analyzed the data using MATLAB using a repeated measure Friedman's ANOVA ($\alpha_{crit} = 0.05$) for the decision times and NASA-TLX workload scores, and Fisher's exact test for the accuracy of the decision. Four comparisons of answer times (global times per display and times for each scenario per display) were performed using a Bonferroni corrected $\alpha_{crit} = 0.05/4 = 0.0125$. A 2x3 Fisher's exact test comparing all three displays for all scenario and a 2x2 Fisher's exact test comparing

the two new displays against the control scenario were performed at a Bonferroni corrected $\alpha_{crit} = 0.05/3 = 0.0167$.

## 5.4 Results

Sixteen nurses (3 males) with a median age of 27.5 years (range 21-53) and 2.75 years of ICU experience (0.5-30 years) participated in the study. One participant had to repeat the training quiz before entering the study, but no one was excluded. The median study duration was 23 min (range 19-39 min).

### 5.4.1 Decision Times

Figure 5.4 shows that nurses were statistically significantly faster in identifying the more critical patient when using the Bar and Clock displays. Median decision times were 11.3 sec for the Bar display, 12.4 sec for the Clock display and 17.2 sec for the Control display. Using the Bar display nurses made their decision 34% faster than the Control display ($p < 1 \cdot 10^{-38}$, $\chi^2 = 75.84$) and 9% faster than with the Clock display ($p = 4.3 \cdot 10^{-3}$, $\chi^2 = 8.17$), while nurses using the Clock display made their decisions 28% faster than with the Control display ($p = 1.3 \cdot 10^{-10}$, $\chi^2 = 41.29$).

Figure 5.5 shows that nurses were statistically significantly faster in identifying the more critical patient for the "Vital Signs Stable," "Past Alarms," "One Alarm," "Ventilator Setting," "Pump Reminder," and "Approaching Alarm" scenarios, when using the Bar display ($p = 9.9 \cdot 10^{-4}$, $1.3 \cdot 10^{-3}$, $4.0 \cdot 10^{-4}$, $6.8 \cdot 10^{-4}$, $1.9 \cdot 10^{-8}$, $8.8 \cdot 10^{-3}$), and "Vital Signs Stable," "One Alarm," and "Pump Reminder" scenarios, when using the Clock display ($p = 3.4 \cdot 10^{-3}$, $1.3 \cdot 10^{-4}$, $7.3 \cdot 10^{-7}$). The only statistically significant difference between the Bar and Clock displays was found for the "Past Alarms" scenario ($p = 4.5 \cdot 10^{-3}$).

### 5.4.2 Decision Accuracy

The overall accuracy was 74.1% with the Clock display, 73.4% with the Bar display, and 66.9% with the Control display. The overall accuracy for the "Past Alarms" scenario was 48% and the overall accuracy of the "Approaching Alarm" scenario was 53%, indicating problems with these two scenarios.

Figure 5.6 shows the accuracy in identifying the more critical patient by scenario:

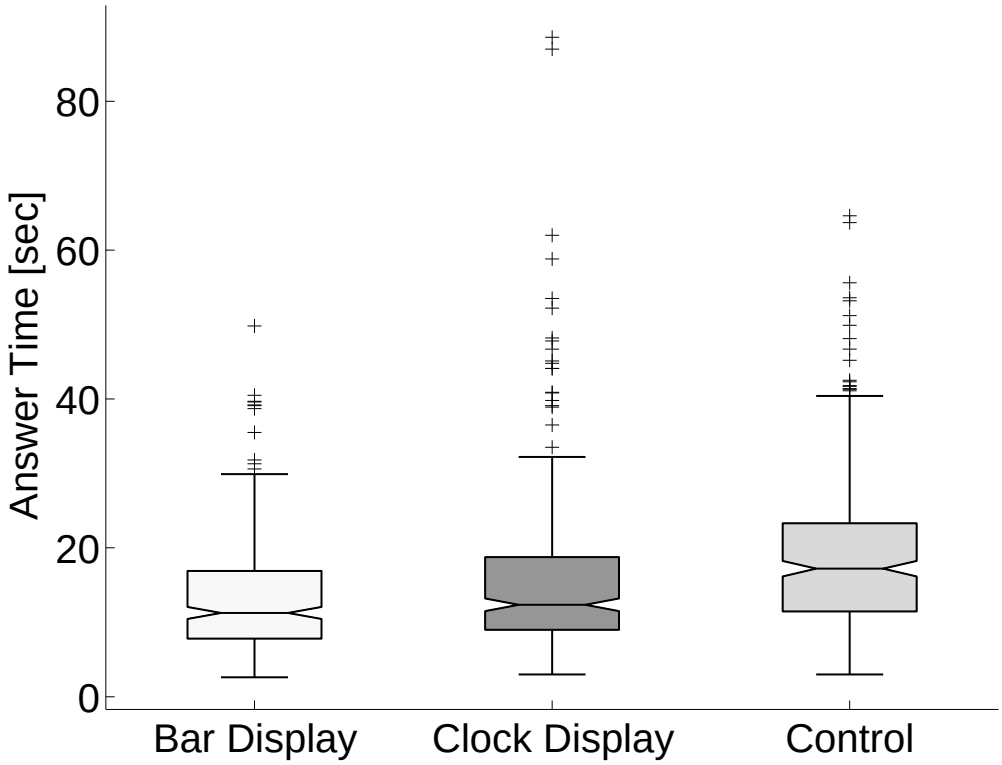


**Figure 5.4**: Answer times for each display. Each icon shows the lowest value, the lower quartile, the median value, the upper quartile and the uppermost value. A plus sign (+) denotes outliers. The improvements of answer times between the Bar display and both other displays as well as the improvement of the Clock display over the Control display were statistically significant.

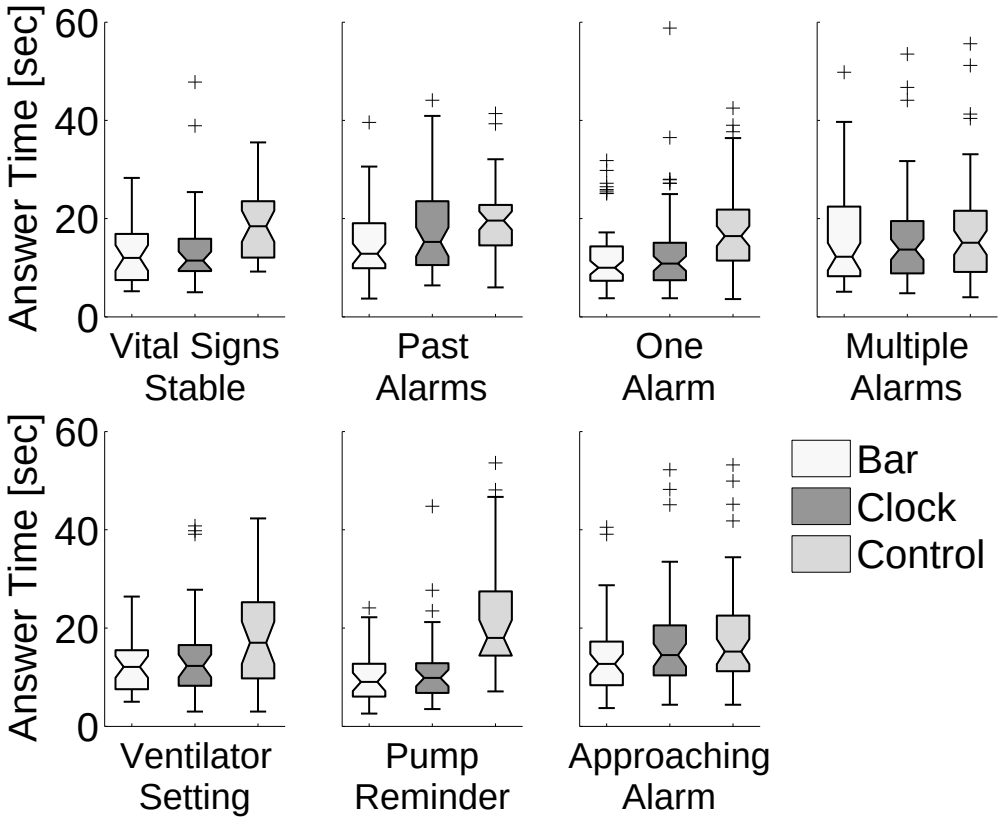


**Figure 5.5**: Answer times for each display, grouped by scenario. Each icon shows the lowest value, the lower quartile, the median value, the upper quartile and the uppermost value. A plus sign (+) denotes outliers. The 7 scenarios (top left to the right) were: 1) Stable patients with no alarms in the past 12 hr, 2) Identify from trend analysis which patient had more past alarms, 3) Identify which patient currently has one alarm, 4) Identify which patient currently has more alarms, 5) Identify which patient requires more ventilator support (higher FiO2 and mechanical MV), 6) Identify which patient has a medication running out in <15 min, and 7) Identify which patient is trending towards an alarm limit.

Statistically significant results in the comparison of all three displays were found for the "Vital Signs Stable" ($p = 4.6 \cdot 10^{-3}$) and "Pump Reminder" ($p = 0.015$) scenarios. We found a significant difference between the Bar display and the Control display for the "Vital Signs Stable" scenario ($p = 2.4 \cdot 10^{-3}$), while the comparison of the Clock and the Control display found a marginally significant difference in the "Pump Reminder" scenario ($p = 0.0164$).

### 5.4.3 Workload Scores and Display Preference

Figure 5.7 shows the self-reported workload scores: For self-reported frustration the two far-view displays performed significantly better than the control scenario ($p = 0.03$). When asked which display they liked best, 9 nurses preferred the traditional display, 3 preferred the Bar display and 4 preferred the Clock display. All 16 nurses mentioned that they liked the syringe icon, showing the infused medication name and the time until this infusion was empty, and requested this information be available outside the patient's room in their unit.

## 5.5 Discussion

Using the two novel far-view displays shown in Figures 5.1 and 5.2, nurses identified the patient in most need of assistance quicker and with greater accuracy. The median decision times were 11.3 sec with the Bar display, 12.4 sec with the Clock display, and 17.2 with the Control display. The new display helped the user more accurately identifying patients with active alarms and nearly empty syringe pumps.

### 5.5.1 Decision Times

The 4.8-5.9 sec reduction in decision time is small, but could lead to an increase in the frequency at which nurses glance at the monitor because the information displayed is easier to see and quickly comprehend. Anesthesiologists glance at their patient monitors 4.3±1.3 times/min[11] whereas nurses in an ICU glance at their patient monitors only 0.3 times/hr to 4.1 times/hr,[1] while both patient types might be comparably critical. Having to enter the patient's room to see the monitor might be one of the reasons this frequency is significantly lower for ICU nurses. The far-view display might increase nurses' vigilance because patient information can be seen from

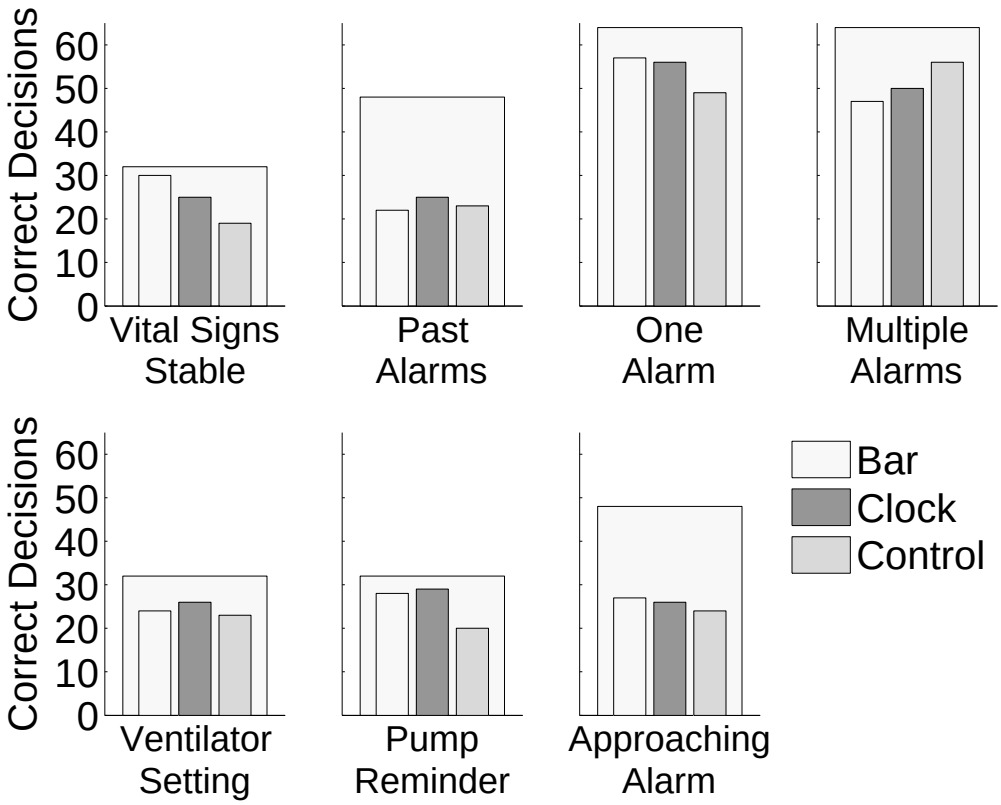


**Figure 5.6**: Answer accuracy for each display, grouped by scenario. The box in the background indicates how often a scenario was experienced during the study (2-4 times per participant), while the smaller bars show the accuracy per scenario and display. The 7 scenarios (top left to the right) were: 1) Stable patients with no alarms in the past 12 hr, 2) Identify from trend analysis which patient had more past alarms, 3) Identify which patient currently has one alarm, 4) Identify which patient currently has more alarms, 5) Identify which patient requires more ventilator support (higher FiO2 and mechanical MV), 6) Identify which patient has a medication running out in <15 min, and 7) Identify which patient is trending towards an alarm limit.

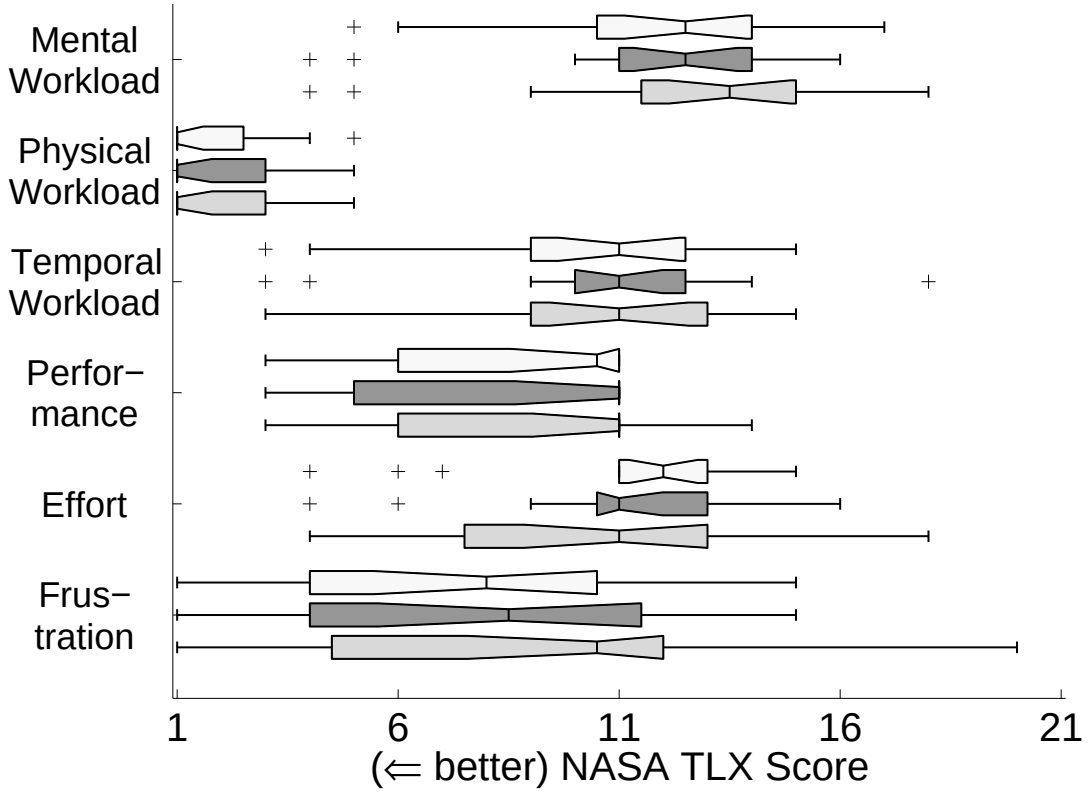


**Figure 5.7**: NASA TLX self-reported workload scores for each display. Light grey indicates the use of the Bar display, dark gray the use of the Clock display and medium gray the use of the Control display. Each icon shows the lowest value, the lower quartile, the median value, the upper quartile and the uppermost value. A plus sign (+) shows outliers. Only the decrease in self-reported frustration was statistically significant.

outside the patient's room and the workload for the triaging task is reduced.

### 5.5.2 Decision Accuracy

Improved accuracy was found in two scenarios: Detecting patients who did not need the provider's attention (with the Bar display) and detecting a nearly empty syringe (with both far-view displays). However, our new displays did not help nurses predict that HR, MAP or SpO2 would reach an alarm threshold within the next 4 min. Miller et al.[12] found that their integrated electronic display helped nurses correctly identify a change in the patient's variables. Adding visual clues that amplify the rate of change[13] could potentially mitigate the observed inattentional blindness problem.[14]

It was surprising that the new displays did not help nurses identify that the patient receiving ventilator support and high FiO2 was more in need of attention than the patient breathing spontaneously at low FiO2. It is most likely due to the fact that nurses are trained to look at SpO2 in context of FiO2 and failed to see the more subtle difference in ventilator support.

Finally, the traditional display had a slightly increased accuracy (but not statistically significant) over the two new far-view displays in triaging multiple alarms ("Multiple Alarms" scenario), which could have been caused by longer trends in the far-view displays distracting the nurse from the current state.

### 5.5.3 Accuracy Difference Between Both Far-View Displays

A statistically significant improvement in accuracy for detecting near-empty syringe pumps was found for the Clock display ($p = 0.016$, $\alpha_{crit} = 0.017$), but not for the Bar display ($p = 0.025$), even though both displays showed identical syringe icons. This is most likely due to the low frequency (10%), this scenario was used. A retro perspective power analysis suggests that 11 additional subjects need to be investigated for the difference to be statistically significant.

With the Bar display nurses were better able to identifying stable patients (94% accuracy, $p = 0.002$) when compared with the Control display (60%). The clock display accuracy (78%, $p = 0.18$) was not significantly better than the Control. This might lead one to assume that the Bar display's linear trend was easier to interpret

and compare than the Clock display's circular trend.

### 5.5.4 Workload Scores and Display Preference

Nurses experienced less frustration with the triage task when they used the new displays. The new displays consolidate the data and reduced the amount of visual scanning required to extract the necessary information. Reducing information overload, providing too much patient data for one nurse to process effectively, is important as this is likely to reduce chances for medical error and thereby eliminate one risk factor for bad patient outcome.[15] However, the nurses felt they performed equally well and had the same mental workload with all three displays. Perhaps the scenarios were not challenging enough to separate workload differences between displays.

### 5.5.5 Comparison with Existing Solutions from the Literature

Information integration can lead to a greater understanding of the patient's state. Integrated information displays have helped anesthesiologists and nurses detect clinical events faster and more accurately and have increased situation awareness.[16–18] Artificial horizons indicate deviations from normal, pie-charts show changes in ventilator settings,[19] polygons and histograms show when vitals signs[20] and blood gas values deviate from normal,[21] and changes in the shape and color of cardiovascular[22] and pulmonary[23] metaphors highlight change. In our Bar display and Clock display deviations from normal may be more salient because regular shapes become distorted and deviations from normal are filled with color.

Anesthesiologists prefer context-specific information[24] and nurses prefer detailed information that highlights cause-and-effect relationships.[12] Our display shows a trend of the patient's minute ventilation plotted alongside a trend showing the minute ventilation provided by the mechanical ventilator, and SpO2 alongside FiO2. Plotting independent and dependent variables together highlights cause-and-effect relationships and improves medical decision making.[25–28]

Multiple vital sign variables can be integrating into a single number to indicate a patient's need for attention.[29, 30] We have added information from infusion pumps or mechanical ventilators. The two far-view displays consolidate information related to the patient's state with information from medical devices in one central location to

facilitate rapid triaging. The new display is not intended to replace the traditional display, but rather augment it by providing summary information when a provider is not at the bedside. A monitor displaying our far-view display might revert to displaying a traditional waveform display when a health care provider is present at the bedside. Alternatively, the far-view display might be shown on a small LCD display by the doorway to the patient's room. Finally, the new displays might find application in remote-ICU monitoring, where physicians monitor multiple patients from a site outside the unit.[31, 32]

### 5.5.6 Limitations

The main limitation of this study was our failure to validate our scenarios by a Delphi method with expert users.[33] In fact, we rechecked and changed the correct answer in one of our "Past Alarms" scenarios when 79% of the volunteers selected that both patients were in equal need of care. Otherwise we relied on the criteria listed in Table 5.1 to designate which of the 2 patients was most in need of nursing care, potentially eliminating more difficult scenarios where the differences between both patients were more subtle. In selecting the patient data to display, we potentially added investigator bias by choosing patients whose need for treatment is better highlighted by features of the far-view displays.

Choosing blue to highlight abnormally low values in our new displays might not have been a good choice, as yellow and red are traditionally used to indicate alarms. However alarms were still highlighted in yellow. The support indicator elements in our display were not intuitive. If users were to receive additional training on use of the support indicators, they might have used them more effectively. One might argue, that we should have evaluated the syringe icon separately, as we expected it to perform well; however, our intent was to evaluate the display in its suggested final form.

A final limitation was the short 3.5-7 min period of training provided for the new displays. The nurses were very familiar with the traditional patient monitor, part of our control display, and might have used it more effectively.

### 5.5.7 Future Work

Information from a work domain analysis[34] or a cognitive task analysis[35] could be used to redesign the display and improve its effectiveness in supporting the triaging task. For a future evaluation we plan to provide the nurses with more training and use think-aloud interviews or eye tracking to provide insight into why the display was effective supporting the triaging task. More complex questions should be used to show display benefits in scenarios, where we currently did not find significant differences. Also higher-level situational awareness levels, with more predictive decision making, should be explored in more detail.[36]

Finally, it would be interesting to evaluate a far-view display with physicians, who are known to make long-term strategic decisions,[37] as these providers could benefit more from the long-term trend component of the display.

### 5.5.8 Conclusion

If implemented in a clinical setting the proposed far-view display might reduce nurses' workload and increase patient safety by enhancing vigilance in three ways:

1. By quickly assessing a patient, nurses could prioritize when to enter a patient's room, and which patient to attend to first. This may allow them to optimize the order in which they perform their tasks

2. Less time required for patient triage could cause this task to be performed more frequently, keeping the provider better informed about the current state of the patient. This might allow initiation of treatment before alarms prompt them to enter the room.

3. The syringe display could improve medication management by identifying empty infusions and the need to order medications without having to enter the patient's room.

## 5.6 Acknowledgments

This research was supported by a research grant from Drägerwerk AG, Lübeck, Germany. The authors would like to thank all participating nurses at the University

of Utah hospital for their support, and Eric Boam for his help with the initial display development.

## 5.7 References

1. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg. 2009 May;108(5):1546–1552.

2. Goodstein L. Discriminative display support for process operators. In: Rasmussen J, Rouse W, editors. Human Detection and Diagnosis of System Failures. Springer; 1981. p. 433–449.

3. Jungk A, Thull B, Hoeft A, Rau G. Evaluation of two new ecological interface approaches for the anesthesia workplace. J Clin Monit Comput. 2000;16(4):243–258.

4. Sanderson PM, Watson MO, Russell WJ. Advanced patient monitoring displays: tools for continuous informing. Anesth Analg. 2005 Jul;101(1):161–168.

5. Waydhas C. Intrahospital transport of critically ill patients. Crit Care. 1999;3(5):83–89.

6. Voigt LP, Pastores SM, Raoof ND, Thaler HT, Halpern NA. Review of a large clinical series: intrahospital transport of critically ill patients: outcomes, timing, and patterns. J Intensive Care Med. 2009 Mar;24(2):108–115.

7. Holtzblatt K, Wendell JB, Wood S. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design. San Francisco, CA: Elsevier/Morgan Kaufmann; 2004.

8. Görges M, Westenskow D, Orr J. Evaluation of a Sodium-Nitroprusside Titration Support System. In: Proceedings of the American Society of Anesthesiologists 2009 Annual Meeting. New Orleans, LA; 2009. p. A403.

9. Korhonen I, Ojaniemi J, Nieminen K, van Gils M, Heikela A, Kari A. Building the IMPROVE Data Library. IEEE Eng Med Biol Mag. 1997 Nov;16(6):25–32.

10. Hart S, Staveland L. Development of NASA-TLX (Task Load Index) Results of Empirical and Theoretical Research. In: Hancock P, Meshkati N, editors. Human Mental Workload. Amsterdam: North Holland Press; 1988. p. 139–183.

11. Liu D, Jenkins SA, Sanderson PM, Fabian P, Russell WJ. Monitoring with head-mounted displays in general anesthesia: a clinical evaluation in the operating room. Anesth Analg. 2010 Apr;110(4):1032–1038.

12. Miller A, Scheinkestel C, Steele C. The effects of clinical information presentation on physicians' and nurses' decision-making in ICUs. Appl Ergon. 2009 Jul;40(4):753–761.

13. Tappan J, Daniels J, Slavin B, Lim J, Brant R, Ansermino J. Visual Cueing with Context Relevant Information for Reducing Change Blindness. J Clin Monit Comput. 2009 Aug;23(4):223–32.

14. Chun MM, Marois R. The dark side of visual attention. Curr Opin Neurobiol. 2002 Apr;12(2):184–189.

15. Donchin Y, Seagull FJ. The hostile environment of the intensive care unit. Curr Opin Crit Care. 2002 Aug;8(4):316–320.

16. Zhang Y, Drews FA, Westenskow DR, Foresti S, Agutter J, Bermudez JC, et al. Effects of Integrated Graphical Displays on Situation Awareness in Anaesthesiology. Cognition, Technology & Work. 2002 Jun;4(2):82–90.

17. Drews FA, Westenskow DR. The right picture is worth a thousand numbers: data displays in anesthesia. Hum Factors. 2006 Spring;48(1):59–71.

18. Görges M, Staggers N. Evaluations of physiological monitoring displays: a systematic review. J Clin Monit Comput. 2008 Jan;22(1):45–66.

19. Liu Y, Osvalder AL. Usability evaluation of a GUI prototype for a ventilator machine. J Clin Monit Comput. 2004 Dec;18(5-6):365–372.

20. Gurushanthaiah K, Weinger MB, Englund CE. Visual display format affects the ability of anesthesiologists to detect acute physiologic changes. A laboratory study employing a clinical display simulator. Anesthesiology. 1995 Dec;83(6):1184–1193.

21. Agutter J, Albert R, Syroid N, Doig A, Johnson K, Westenskow D. Arterial Blood Gas Visualization for Critical Care Clinicians. In: Proceedings of the Annual Meeting of the Society for Technology in Anesthesiology. San Diego, CA; 2006. .

22. Agutter J, Drews F, Syroid N, Westneskow D, Albert R, Strayer D, et al. Evaluation of graphic cardiovascular display in a high-fidelity simulator. Anesth Analg. 2003 Nov;97(5):1403–1413.

23. Wachter SB, Johnson K, Albert R, Syroid N, Drews F, Westenskow D. The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. J Am Med Inform Assoc. 2006 Nov;13(6):635–642. Evaluation Studies.

24. Nazir T, Beatty PC. Anaesthetists' attitudes to monitoring instrument design options. Br J Anaesth. 2000 Nov;85(5):781–784.

25. Blike GT, Surgenor SD, Whalen K, Jensen J. Specific elements of a new hemodynamics display improves the performance of anesthesiologists. J Clin Monit Comput. 2000;16(7):485–491. Clinical Trial.

26. Cole WG, Stewart JG. Human performance evaluation of a metaphor graphic display for respiratory data. Methods Inf Med. 1994 Oct;33(4):390–396.

27. Effken JA, Kim NG, Shaw RE. Making the constraints visible: testing the ecological approach to interface design. Ergonomics. 1997 Jan;40(1):1–27.

28. Effken J, Loeb R, Kang Y, Lin Z. Clinical information displays to improve ICU outcomes. Int J Med Inform. 2008 Jul;.

29. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth. 2006 Jul;97(1):64–68.

30. Taft A, Ronen M, Epps C, Waugh J, Wales R. A Novel Integrated Pulmonary Index (IPI) Quantifies Heart Rate, Etco2, Respiratory Rate and Spo2. In: Anesthesiology. 109; 2008. p. A1682.

31. Breslow MJ, Rosenfeld BA, Doerfler M, Burke G, Yates G, Stone DJ, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: an alternative paradigm for intensivist staffing. Crit Care Med. 2004 Jan;32(1):31–38.

32. Lilly CM, Thomas EJ. Tele-ICU: Experience To Date. J Intensive Care Med. 2010 Jan-Feb;25(1):16–22.

33. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 2001 Apr;38(2):195–200.

34. Burns CM, Hajdukiewicz JR. Ecological Interface Design. Boca Raton, FL: CRC Press; 2004.

35. Crandall B, Klein G, Hoffman RR. Working Minds: A Practitioner's Guide to Cognitive Task Analysis. Cambridge, MA: The MIT Press; 2006.

36. Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. Hum Factors. 1995 Mar;37(1):32–64.

37. Miller A, Sanderson P. Clinical information use by medical and nursing staff in the ICU. In: Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society. Orlando, FL; 2005. p. 984–988.

# CHAPTER 6

# CONCLUSION

## 6.1 Central Theme

The central theme of this dissertation, titled "Signal processing, human factors, and modelling to support intensive care unit bedside care," is to advance technology and improve patient care.

### 6.1.1 Four Manuscripts

It consisted of four manuscripts: 1) a review of previous medical display evaluations, providing insight into solutions that have worked in the past; 2) a study on reducing false alarms and increasing the usefulness of the remaining alarms by introducing alarm delays and detecting alarm context, such as suctioning automatically silencing ventilator alarms; 3) a study of simplifying the frequent but complicated task of titrating of vasoactive medications by providing a titration support tool that predicts blood pressure changes 5 min into the future; and 4) a study on supporting the triage of unfamiliar patients by introducing a far-view display, which incorporates information from previously disparate devices and presents trend and alarm information at one easy to scan and interpret location.

### 6.1.2 Contribution of the Four Parts to the Central Theme

The four manuscripts tie together through their focus on reducing nurses' workload and improving nurses' situational awareness, thereby improving medical decision making, reducing chances for medical errors (see Figure 6.1). The review laid the foundation of the work by identifying previous approaches that have worked in the past and highlighting the need to focus on nurses. The observational study not only provided a simple suggestion for reducing unnecessary alarms by introducing a short alarm delay, but also identified opportunities for improvements: a) introducing a

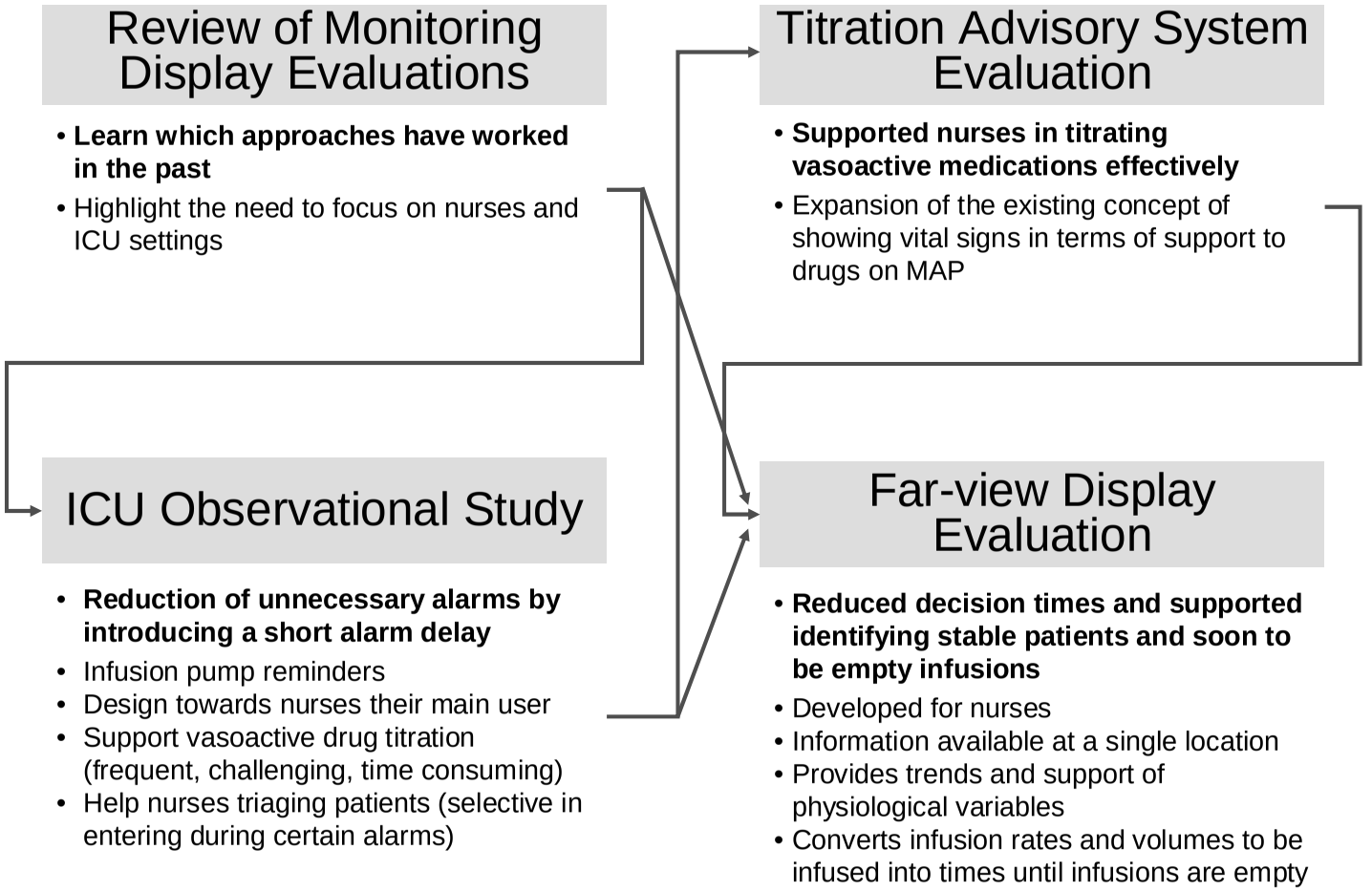


**Figure 6.1**: Manuscript summaries and how they tie together

reminder of soon to be empty infusion pumps; b) designing monitoring devices with nurses, their most frequent users, in mind; c) supporting vasoactive drug titration, a common, challenging, and time consuming task; and d) helping nurses decide when to enter a patient's room, as we found that nurses most frequently entered the room to respond to an effective alarm, which caused an action to be performed, rather than to just silence the alarm. The evaluation of the titration advisory system not only supported nurses in titrating vasoactive medications effectively, an area of potential improvement identified in the observational study, but it also expanded the existing concept of showing vital signs in terms of support, such as the blood oxygen saturation in terms of inspired oxygen fraction and minute ventilation, to medication support on arterial blood pressure. Finally, the far-view display evaluation incorporated improvement opportunities previously identified, such as infusion pump reminders and support indicators. Its development also focused on nurses as their main users and acknowledged nurses' information requirements by: a) combining information from multiple devices at single location visible without having to enter the patient's room, b) providing trends, and c) converting infusion rates and volumes to be infused into times infusions were empty.

## 6.2 Summary and Conclusions

### 6.2.1 Review of Physiologic Monitoring Display Evaluations

The review of evaluation studies for physiologic monitoring displays (Chapter 2) found that study participants detected an adverse event or made a diagnosis or decision more rapidly in 17 of 30 studies, were more accurate in a clinical decision or diagnosis in 12 of 18 studies, and showed decreased perceived workload in 3 of 7 studies. These results demonstrate that improved displays have high potential in improving medical decision making and reducing clinicians' workload and chances for medical error. Finally, the review highlighted the need to use study participants other than anesthesiologists, who participated in 18 of 30 studies, and focus especially on nurses, who participated in only 5 of the 30 studies, even though they are known to conduct a majority of patient care tasks.

### 6.2.2 Alarm Reductions Using Delays and Clinical Context

The study in Chapter 3 recorded time-stamped information of alarms and the presence of health care team members in the patient room for 200 hr. It found that many alarms are transient (canceling themselves when the offending condition passes) and/or associated with common patient care tasks, such as suctioning the airway or manipulating the patient. Introducing a 19 sec alarm delay and automatically detecting suctioning, repositioning, oral care, and washing could reduce the number of ineffective and ignored alarms from 934 to 274—a 71% reduction. More reliable alarms could elicit more timely responses, reduce workload, reduce noise pollution and potentially improve patient safety. Finally, it was observed that nursing staff intentionally entered a smaller infusate volume than was available so that the infusion pump alarm reminded them when the pump was nearly empty—here, an automated reminder could eliminate the number of duplicate alarms (the reminder and one or more additional alarms when the infusion is really empty, depending on whether the nurse correctly entered the remaining volume and did not want to waste any).

### 6.2.3 Titration Advisory System with Patient Specific Sensitivity Identification

Titrating vasoactive drug infusion rates is a frequent but challenging nursing task, because of significant variations in patients' sensitivities and delays between changes in infusion rates and observed changes in blood pressure exits. A potential solution to this problem is providing an open-loop advisory system, which predicts mean arterial pressure (MAP) 5 min into the future (Chapter 4). The proposed advisory system shortened the median time required by nurses to reach the desired MAP from 10.2 to 4.1 min (a 60% reduction), decreased the median number of infusion rate changes from 6 to 4 (a 33% reduction), and resulted in a significant reduction of mental workload and effort. Identifying an individual patient's sensitivity, instead of predicting MAP based on an "average" patient's sensitivity, improved the accuracy of the prediction by 75% for sodium-nitroprusside, 82% for dopamine, and 52% for dobutamine. By predicting and displaying the expected blood pressure 5 min in the future, the advisory system helped nurses titrate faster and reduced their perceived

workload and might improve patient safety.

### 6.2.4 Intensive Care Unit Far-View Display Supporting Triaging Tasks

Triaging patients and deciding which of 2 patients to attend to first is a task nurses perform multiple times each hr. This task requires integration of information from three disparate devices—the patient monitor, the ventilator, and the infusion pumps—and generally requires entering the patient's room (as these devices might not be readable from the door), thereby potentially disturbing the patient's rest. This task might be simplified by introducing a far-view display (Chapter 5), which incorporates information from the three mentioned devices and presents trend and alarm information at one easy to scan and interpret location. Additionally, infusion pump reminders, a problem identified in the observational study, and therapy support indicators, a modified approach to the prediction system introduced in the titration advisor paper, were included. Finally, we designed the novel display for nurses as the main users, incorporating lessons learned from the literature review (Chapter 2), which identified nurses as a previously understudied population specifically, and the alarm observation study (Chapter 3), which demonstrated that nurses are the main monitor users. Using the two proposed far-view displays, nurses more accurately identified stable patients and nearly empty syringe pumps. Median decision times improved from 17.2 sec for the control display to 11.3 sec for the bar display and 12.4 sec for the clock display—a 33-35% reduction. By graphically integrating disparate information, the far-view display might reduce nurses' workload, improve nurses' decision making, and increase patient safety by allowing nurses to more quickly detect patients in immediate need of attention.

## 6.3 Impact

The presented work demonstrates two improvements towards nurses' situational awareness: a) reductions of nurses' workload and chances for medical error by introducing the titration advisory system for vasoactive drug infusions and reducing the number of irrelevant and ignored alarms; and b) improvements in medical decision making by introducing a far-view display that supports the triaging of patients.

The combination of improvements in signal processing (alarm reduction strategies in Chapter 3 and titration advisor in Chapter 4), human factors (titration advisor in Chapter 4 and far-view display in Chapter 5), and modeling (titration advisor in Chapter 4) to support ICU bedside care have potential for reducing the number of medical errors by improving nurses' situational awareness, helping them make more accurate and faster decisions, and reducing their workload. Finally, reducing irrelevant and ignored alarms could reduce ICU noise levels and patient interruptions, which might have beneficial effects on patient outcome.[1, 2]

## 6.4 Future Work

### 6.4.1 Review of Physiologic Monitoring Display Evaluations

As the review, which formed Chapter 2 of this dissertation, was published a couple of years ago, an update reviewing evaluation studies performed in the last 3.5 years could lead to insights into recent improvements in evaluation techniques. Such an update should also focus on explaining how frameworks may be used or designed by the researcher performing an evaluation, and how information obtained through work domain analyses can be included in display design.

### 6.4.2 Alarm Reductions Using Delays and Clinical Context

Using the data collected in the alarm study, which formed chapter 3 of this dissertation, one could reanalyze the effectiveness of past alarm systems in bringing a provider into the room, e.g., it was observed that of the 524 alarms that sounded when no healthcare provider was present in the patient's room, a healthcare provider entered during, or within 2 min, of these alarms only in 111 and 180 episodes, respectively. This suggests that alarms are not felt to properly reflect the patient condition and are thus not recognized as helpful. With noise in the ICU having a detrimental effect on patients' sleep and ICU outcome, auditory alarms should sound outside the patient's room or be transmitted directly to the responsible healthcare provider if none is present at the bedside. Here, a personalized alarm system communicating patient alarms directly to the nurse when he or she is not in the respective patient's room could be designed and evaluated.

Other ways of reanalyzing the existing data include looking at the relative importance and optimal duration of delays for individual parameters, sensors or devices and the analysis of repeated alarm efficacy, e.g., alarms for infusion pumps and feeding pumps could be exempt from the delay as their alarms do not terminate themselves, or TV, SpO2 and ABP could have different delays based on the desired reduction in ineffective and ignored alarms. (For a 50% reduction TV alarm delays only need to be 9 sec, while ABP delays need to be 22 sec.)

Finally, our study could be expanded in two ways: a) observing alarms during different times of the day and in other units, and b) more importantly, repeating the study in a ICU with modern equipment using current alarm delay settings and then again after optimizing the alarm delays as proposed in the evaluation to demonstrate the "real" clinical benefit suggested by our observations.

### 6.4.3 Titration Advisory System with Patient Specific Sensitivity Identification

The improvements observed in the titration advisor evaluation in Chapter 4 should be verified in a study using patients where nurses would manually change the infusion rates to elicit small MAP changes and use their best clinical judgment to make infusion rate changes based on our predictions. Next, the process of making small changes in the infusion rate, which is required for the identification, could be automated.

Another area for future work is the expansion of this method to other drugs (e.g., norepinephrine[3, 4]), and the expansion of the method to predict changes in other vital signs (e.g., heart rate and cardiac output). Animal experiments using swine could be used to compare the performance of the proposed sensitivity identification method using medications such as dopamine, dobutamine, epinephrine, norepinephrine. Exploring combinations thereof, which are frequently used in the clinical setting,[5, 6] should also be considered. Another area of interest could be the administration of small fluid boluses to identify the volume response of the patient and whether the patient would benefit from additional volume increase.

An interesting expansion of this project would be the application of patient specific sensitivity identification for the use in anesthesiology drug effect displays.[7, 8] Before induction and during stable phases of anesthetic maintenance the patient could be

challenged using changes in sedative and analgesic drug infusion rates or anesthetic vapor concentrations with measurements like processed EEG, heart rate variability or skin conductance serving as surrogates for anesthetic depth measurements and perceived pain levels respectively. A similar concept for mechanical ventilation could be performed where FiO2, PEEP, and/or pressure support would be slightly reduced to identify a change in the patients' SpO2, thereby identifying whether the patient could be adequately ventilated with less mechanical ventilator support.

### 6.4.4 Intensive Care Unit Far-View Display Supporting Triaging Tasks

The improvements recorded in the far-view display evaluation, demonstrated in Chapter 5, should be verified in clinical practice, with the number of "unnecessary" patient-room entries counted. However, the problem with nurses missing trends towards an alarm with the two new displays needs to be addressed first. It could be mitigated by combining the display with change indicators suggested by Tappan et al.,[9] after which the efficacy of the improved far-view display should be evaluated again using more- and less-obvious scenarios. Consequently, a display prototype that interfaces the cardiac patient monitor, mechanical ventilator and infusion pumps could be placed in the window of a patient's room, allowing health care providers to get more familiar with the display and learn its use in real patients. This learning experience might mitigate the lack of training and lack of experience observed in the present work, potentially leading to better performance in future evaluations.

The far-view display should also be combined with Sven Koch's close-view nursing display,[10] in which the display content would depend on provider presence detection technology, such as radio-frequency identification (RFID)[11–13] or ultrasound identification (USID).[14] This combined display could then be evaluated using a human patient–simulator study looking at tasks that require entering the room and interacting with the display (such as medication rate changes or adding drugs to be infused) and tasks performed from the doorway, (such as triaging a patient or identifying alarm sources). An interesting and potentially worth-while side project would be the evaluation of effectiveness of using the clock display as an alternative to the current paper information in communicating a patients' performance in the ICU

during change of shift reports, where the patient's current nurse transfers care to the next nurse.

Finally, the improvements in decision performance should also be verified with physicians, who are more used to seeing trend information and making long term decisions, rather than nurses, who make tactical decisions for the present. Here it could be interesting to see how healthcare providers with different amounts of ICU experience and different priorities, such as pulmonary fellow physicians, and cardiology fellow physicians perform with the two new far-view displays.

## 6.5 References


1. Christensen M. Noise levels in a general intensive care unit: a descriptive study. Nurs Crit Care. 2007 Jul;12(4):188–197.

2. Gabor JY, Cooper AB, Hanly PJ. Sleep disruption in the intensive care unit. Curr Opin Crit Care. 2001 Feb;7(1):21–27.

3. Beloeil H, Mazoit JX, Benhamou D, Duranteau J. Norepinephrine kinetics and dynamics in septic shock and trauma patients. Br J Anaesth. 2005 Dec;95(6):782–788.

4. Johnston AJ, Steiner LA, O'Connell M, Chatfield DA, Gupta AK, Menon DK. Pharmacokinetics and pharmacodynamics of dopamine and norepinephrine in critically ill head-injured patients. Intensive Care Med. 2004 Jan;30(1):45–50.

5. Leone M, Martin C. Vasopressor use in septic shock: an update. Curr Opin Anaesthesiol. 2008 Apr;21(2):141–147.

6. Overgaard CB, Dzavik V. Inotropes and vasopressors: review of physiology and clinical use in cardiovascular disease. Circulation. 2008 Sep;118(10):1047–1056.

7. Syroid ND, Agutter J, Drews FA, Westenskow DR, Albert RW, Bermudez JC, et al. Development and evaluation of a graphical anesthesia drug display. Anesthesiology. 2002 Mar;96(3):565–575.

8. Drews FA, Syroid N, Agutter J, Strayer DL, Westenskow DR. Drug delivery as control task: improving performance in a common anesthetic task. Hum Factors. 2006 Spring;48(1):85–94.

9. Tappan JM, Daniels J, Slavin B, Lim J, Brant R, Ansermino JM. Visual Cueing with Context Relevant Information for Reducing Change Blindness. J Clin Monit Comput. 2009 Aug;23(4):223–32.

10. Koch SH, Staggers N, Weir CR, Agutter J, Liu D, Westenskow DR. Integrated Information Displays for ICU Nurses: Field Observations, Display Design, and Display Evaluation. In: Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society. 429. San Francisco, CA; 2010. .

11. Chen CI, Liu CY, Li YC, Chao CC, Liu CT, Chen CF, et al. Pervasive Observation Medicine: The Application of RFID to Improve Patient Safety in Observation Unit of Hospital Emergency Department. Stud Health Technol Inform. 2005;116:311–315.

12. Dempsey M. Fundamentals of Indoor Positioning Systems. BI&T. 2004 May;p. 293–297.

13. Fry EA, Lenert LA. MASCAL: RFID tracking of patients, staff and equipment to enhance hospital response to mass casualty events. AMIA Annu Symp Proc. 2005;p. 261–265.

14. Holm S, Holm R, Hovind O. System and method for position determination of objects. Sonitor Technologies AS; 2009. United States patent US 7535796.